Structure, function and genetic diversity of glucosyltransferase IV (GtrIV) of *Shigella flexneri*

Anesh Nair

August 2012

A thesis submitted for the degree of Doctor of Philosophy

at the Australian National University

Australian National University

Bacterial and Bacteriophage Genetics & Vaccine Development Laboratory

Research School of Biology

The Australian National University

Preface

This thesis is comprised of a literature review (Chapter 1), a methods chapter (Chapter 2), four results chapters (Chapter 3-6) and a general discussion (Chapter 7).

Declaration

I declare that the results presented within this thesis are, except otherwise acknowledged, my own original work while enrolled as a PhD candidate at the Australian National University (2008-2012)

Anesh Nair

Associate Professor Naresh Verma

Acknowledgments

First and foremost, I would like to thank my supervisor and mentor A/Prof Naresh Verma for his patience and guidance. Thank you for taking me under your wing all those years ago and encouraging me throughout the way. Your door was always open when I needed your advice and expertise. Also, thanks for improving my badminton skills! I also want to thank the members of my supervisory panel Dr Gwen Allison and Prof. Stefan Broer for their valuable input during the course of my PhD project

A big thank you must also go out to all the past and present members of the Verma lab I had the privilege of working and mingling with especially Rob, Bec, Farah, Angeline, Roy, Swee Seong, Richa, Divya, Aruna, Dave, Aviel, Zarshis and Tony. You people have helped keep me sane by putting up with my various antics and laughing at my bad jokes! I want to single out my 3 'brothers' Harris, Pravin and Vinson for special mention. To Harris and Pravin, those inspired 'coffee and cake' sessions were something I looked forward to everyday. You guys are two of the best buddies anyone can have! To Vinson, thanks for all your help with anything IT related. You are a life saver and a true friend.

I want thank my wonderful family for all their support and patience. To my parents, grandmother and my brother, I DID IT! To my beautiful son Ethan, Acha will be spending more time with you now! Finally, to the love of my life, my wife Nora who has put up with me more than anyone else. Thanks for your undying support and love and also for all those wonderful meals you whipped up to lift my spirits. You, more than anyone, gave me strength for this. Above all else, thank you God.

Publications

Nair, A., Korres, H. and Verma, N. K. (2011). Topological chracterisation and identification of critical domains within glucosyltransferase IV (GtrIV) of *Shgella flexneri*. *BMC Biochemistry*. **12**: 67-81.

Nair, A. and Verma, N. K. (2012 - In Preparation). Acidic residues that are critical to glucosyltransferase IV (GtrIV) function in *Shigella flexneri*.

Abstract

One of the required virulence determinants for *Shigella flexneri* pathogenesis is the lipopolysaccharide and in particular the O-antigen component. There are 15 known *S. flexneri* serotypes, which differ in their virulence, prevalence, distribution and the nature of their O-antigens. O-antigen modification occurs via glucosylation, which is a process that is brought about by the addition of glucosyl groups to one or more of the O-antigen sugars. The three bacteriophage genes *gtrA, gtrB* and $gtr_{(type)}$ are responsible for O-antigen glucosylation in *Shigella flexneri*. Both *gtrA* and *gtrB* have been demonstrated to be highly conserved and interchangeable among serotypes while $gtr_{(type)}$ was found to be specific to each serotype, leading to the hypothesis that the $Gtr_{(type)}$ proteins are responsible for attaching glucosyl groups to the O-antigen in a site- and serotype- specific manner. Based on the confirmed topologies of GtrI, GtrII and GtrV, such interaction and attachment of the glucosyl groups to the O-antigen has been postulated to occur in the periplasm.

In this study, the topology of GtrIV, a *S. flexneri* glucosyltransferase responsible for converting serotype Y to serotype 4a, was experimentally determined by creating different fusions between GtrIV and a dual-reporter protein, PhoA/LacZ. This study shows that GtrIV consists of 8 transmembrane helices, 2 large periplasmic loops, 2 small cytoplasmic N- and C-terminal ends and a re-entrant loop that occurs between transmembrane helices III and IV. Though this topology differs from that of GtrI, GtrII, GtrV and GtrX, it is similar to that of GtrIc.

Based on the newly elucidated topology of GtrIV, the roles of the N-terminal and C-terminal periplasmic loops of GtrIV (loop No. 2 and loop No. 6) were investigated by carrying out a series of loop deletions. By sequentially deleting loop segments in loop No. 6, the presence of a potential catalytic site located between resides D260 to W269 was hypothesised. This study also investigated the roles of conserved or specific functions of the two periplasmic loops of GtrIV by creating loop swap chimeras between GtrIV and its closest structural homologue, GtrIc. The resulting hybrids lost their native function and were unable to substitute function to the other protein thus, signifying the importance of both loops in GtrIV function.

A total of 20 negatively charged amino acids that occur throughout GtrIV were mutated to alanine to investigate their role in GtrIV function. Three of these acidic residues located in the periplasmic loop No. 6, D261, E262 and D267, when mutated in tandem, were found to abolish function of GtrIV. They are thought to be involved in the interaction with the donor and acceptor substrates that help confer serotype specificity in O-antigen glucosylation.

This study also looked at 28 wildtype strains isolated from Japan, Bangladesh and Vietnam that were originally typed as serotype 4a, 4b or just 4 and were subjected to a series of slide agglutination experiments to confirm their serotype. A number of strains had atypical agglutination patterns, highlighting the possibility that they are potentially new serotypes or subserotypes. The GtrIV sequences obtained from these strains revealed the presence of point mutations within several strains, although we were not able to identify mutations that were detrimental to the function of GtrIV. Southern blots carried out using DIG-labelled *gtrIV*

revealed that two different serotype 4a strains, with the possibility of one of them being originally a 4b strain, have been circulating within these populations. Taken together, this study has provided valuable insight into the possibly unique mode of action of GtrIV and has provided grounds for further investigation into the evolution of one of the most important human pathogens.

Abbreviations

aa	amino acid
Amp	ampicillin
att	attachment site
BCA	bicinchoninic acid
bp	base pairs
Cm	chloramphenicol
C-terminal	carboxyl terminus
DI	dual indicator
DIG	digoxigenin
D-MALT	diffuse mucosa associated lymphoid tissue
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
EDTA	ethylene diamine tetracetic acid
Exo III	exonuclease III
Gal	galactose
Glu	glucose
Glu*N*Ac	N-acetyl glucosamine
Gtr	glucosyltransferase
GT	glycosyltrasferase
ics	intracellular spread gene
IFN	interferon
Ig	immunoglobulin

IL	interleukin
int	integrase
Ipa	invasion plasmid antigen
IPTG	isopropylthiogalactoside
kb	kilobase
Kan	kanamycin
LacZ	β-galactosidase
LB	Luria-Bertani medium
LPS	lipopolysaccharide
MAC	membrane attack complex
MASF IV	monoclonal antibody specific to *S. flexneri* serotype 4a and 4b
MASF Ic	monoclonal antibody specific to *S. flexneri* serotype 1c
Mb	megabase
M cells	microfold cells
MCS	multiple cloning site
mM	millimolar
Mxi	membrane excretion proteins
NAR	Normalized Activity Ratio
NCTC	National Collection of Type Cultures (UK)
N-terminal	amino terminal
Oac	O-acetyltransferase
orf	open reading frame
PAGE	polyacrylamide gel electrophoresis

PAI	pathogenicity island
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PhoA	alkaline phosphatase
PMN	polymorphonuclear cell
RE	restriction enzyme
Red-Gal	6-chloro-3-indolyl-β-D-galactoside
Rha	rhamnose
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
SfII	serotype-converting phage containing *gtrII* locus
SfV	serotype-converting phage containing *gtrV* locus
SfX	serotype-converting phage containing *gtrX* locus
Sf6	serotype-converting phage containing *oac* gene
SNP	single nucleotide polymorphism
Spa	surface presentation antigens
SSC	Salt-Sodium Citrate buffer
SSPE	Salt-Sodium Phosphate–EDTA buffer
TBE	Tris-Borate-EDTA buffer
TE	Tris-EDTA buffer
TM	transmembrane helix
T_m	melting temperature
UndP	undecaprenol phosphate

UV	ultraviolet radiation
V	volts
WHO	World Health Organisation
Xphos	5-bromo-4-chloro-3-indolyl phosphate
2-ME	2 Mercaptoethanol

Table of Contents

Chapter 1

Introduction & Literature Review

1.1 Shigellosis

Shigella is the major causative agent of bacillary dysentery (shigellosis), which is characterised by fever, abdominal cramps and bloody mucopurulent stools (Kotloff *et al.*, 1999). *Shigella* belongs to the Gram-negative bacterial family *Enterobacteriaceae*, which contains facultatively anaerobic straight rods with simple nutritional requirements, and includes other important pathogenic genera such as *Escherichia* and *Salmonella* (Sansonetti, 2001; Niyogi, 2005; Zhu *et al.*, 2009). If they were classified today, the *Shigella* species and *E. coli* would comprise one species, since they are very closely related genetically (Jin *et al.*, 2002; Wei *et al.*, 2003; Christopher *et al.*, 2010). There are four major serogroups in this genus namely, *Shigella flexneri, Shigella sonnei, Shigella boydii* and *Shigella dysenteriae* (Hale, 1991; Phalipon & Sansonetti, 2007; Barnoy *et al.*, 2010). These serogroups differ greatly in their O-antigen structures, their prevalence, their diversity and the severity of the infection caused (Yang *et al.*, 2005).

A review of literature between 1966 and 1997 estimated that the annual number of *Shigella* cases was 163.2 million in developing countries, with 1.1 million deaths (Kotloff *et al.*, 1999). In industrialized countries, the number of cases was much less, with only 1.5 million cases of *Shigella* cases annually (Kotloff *et al.*, 1999). It was also estimated that 69% of all *Shigella* cases and 61% of all deaths due to the disease occurred in children under the age of five (Farshad *et al.*, 2006). A more recent study carried out by Ram *et al.* (2008) found that the actual incidence of shigellosis cannot yet be measured due to the lack of epidemiological studies in low–human development index (HDI) countries such as those in sub-Saharan Africa. Due to the

limited access to clean water supplies and sanitation facilities, low-HDI countries are expected to have a higher incidence of shigellosis in comparison to those countries that have been documented in the studies (Ram *et al.,* 2008). *Shigella* infections are the second most common cause of traveller's diarrhoea (Svennerholm and Steele, 2004).

The different serogroups of *Shigella* vary greatly in their prevalence. Previous epidemiological studies indicate that the distribution of the different serogroups is heavily dependent on socioeconomic factors (Opintan & Newman, 2007). It was further found that *S. flexneri* was the main serogroup responsible for causing shigellosis in developing countries, with a median percentage isolate of 60% (Kotloff *et al.,* 1999). *S. sonnei* had the second highest median percentage isolate of 15%, followed by *S. dysenteriae* and *S. boydii*, with each having a 6% median isolate percentage (Kotloff *et al.,* 1999; Ram *et al.,* 2008). In industrialized countries, the median percentage of isolated serogroups differed greatly. *S. sonnei* had the highest percentage (77%), followed by *S. flexneri* (16%), *S. boydii* (2%) and *S. dysenteriae* (1%) (Kotloff *et al.*, 1999; Ram *et al.*, 2008).

The clinical signs caused by *Shigella* can range from mild diarrhoea to severe dysentery. Dysentery is diarrhoea in which blood and mucus are present in the stools. Some other common symptoms include fever, nausea, vomiting, intestinal cramps, abdominal pain and seizures (Niyogi, 2005; Sansonetti, 2001). Inflammation in the colon and epithelial destruction gives rise to this clinical manifestation in shigellosis (Taylor *et al.,* 1986; Jennison *et al.,* 2006). In rare cases, the *Shigella* bacteria can invade the blood, lungs or brain leading to serious and often fatal complications such as haemolytic uremic syndrome, Gram-negative septicaemia, acute

hypoglycaemia, megacolon, or seizures (Bennish, 1991; Sansonetti, 2001; Daneman & Rescigno, 2009; Christopher *et al.*, 2010; Butler, 2012). Furthermore, salt and water fluctuations due to diarrhoea and dysentery can lead to kidney malfunction and failure (Christopher *et al.*, 2010). In very severe cases, shigellosis can even lead to death. This especially occurs in young children, infants below five years of age and other immunodeficient individuals and is commonly accompanied by fever, abdominal pain and anorexia (Bennish, 1991; Ferreccio *et al.*, 1991; Kosek *et al.*, 2010; Wong *et al.*, 2010).

Shigellosis is transmitted mainly via the faecal-oral route. The consumption of contaminated water or food is the most common means of disease acquisition. The transmission of the disease is also enhanced by poor hygiene, malnutrition and the lack of medical treatment (Al Jarousha *et al.*, 2010; Vu Nguyen *et al.*, 2006). Shigellosis is highly contagious, as it has a low infective dose, whereby only 10-100 bacterial cells are needed to establish the infection (DuPont *et al.*, 1989; Phalipon & Sansonetti, 2007; Sousa *et al.*, 2010). As a result of the highly infectious nature of *Shigella*, the disease can also be spread by person-to-person contact. This mode of transmission is most frequently encountered in day-care centres for children and other institutions (in both developing and industrialised nations) where personal hygiene standards are low (Pickering *et al.*, 1981; Mahoney *et al.*, 1993).

1.2 Pathogenesis of *Shigella flexneri*

An understanding of the pathogenesis of *Shigella* is vital for developing vaccine candidates. Elucidation of the pathogenesis of *Shigella* has been complicated by the absence of a good animal model. Only higher primates develop a similar illness to humans, and practical

considerations limit their use. The mouse pulmonary model (Mallet *et al.*, 1993), guinea pig eye and colitis models (Sereny, 1957; Shim *et al.,* 2007; Kweon M. N., 2008), and rabbit intestinal loop model (Wassef *et al.*, 1989) have been used to study some aspects of *Shigella* pathogenesis, but caution must be applied when extrapolating these findings to humans.

Through the up regulation of acid resistance genes, *Shigella* survive their passage through the acidic environment of the stomach (Small *et al.*, 1994; Espina *et al.*, 2006). The pathogen then employs three methods to cross the colonic epithelial layer in order to gain access to the basolateral sides of the mucosal cells for invasion.

Firstly, tight junction proteins can be manipulated to allow the pathogens to reach the basal side of the epithelia layer (Sakaguchi *et al.*, 2002; Jennison & Verma, 2004, Kaminski & Oaks, 2009). *S. flexneri* also exploits the highly endocytic M cells (microfold cells) of the mucosal-associated lymphoid tissue, which mediate the translocation of lumenal antigen and microorganisms across the epithelial barrier to the inductive sites of mucosal immunity (Jepson and Clark, 1998; Ogawa & Sasakawa, 2006; Kaminski & Oaks, 2009). The bacteria are able to survive the killing mechanisms of the macrophages by an IpaB-mediated (protein effector) lysis of the phagocytic vacuole and initiates macrophage apoptosis that causes the release of cytokines interleukin-1β (IL-1β) and interleukin-18 (IL-18) (Barzu *et al.,* 1996; Chen *et al.*, 1996; Neutra MR, 1999; Espina *et al.,* 2006; Ogawa & Sasakawa, 2006; Ashida *et al.*, 2011). These cytokines cause the charactertistic intestinal inflammation of shigellosis (Ashida *et al.*, 2011). *Shigella* is then able to invade the epithelial cell from the basolateral side, triggering a strong inflammatory response that leads to the secretion of the chemokine interleuine-8 (IL-8) that helps recruit

polymorphonuclear (PMN) cells to the site of infection by transmigrating through the epithelial lining to reach luminal bacteria (Girardin *et al.,* 2003; Singer & Sansonetti, 2004; Schroeder and Hilbi, 2008; Ashida *et al.*, 2011). This disruption of the epithelium allows further shigellae to cross the colonic epithelial layer (Perdomo *et al.*, 1994; Beatty and Sansonetti, 1997). The three mechanisms by which *Shigella* can cross the colonic epithelial layer are illustrated in Figure 1.1.

To counteract the host's defences, *Shigella* express a needle-like type III secretion system (T3SS) that consists of cytoplasmic, transmembrane, and extracellular domains (Deane *et al.,* 2006) (Figure 1.2). Physical contact with host cells initiates secretion of effector proteins and leads to formation of a pore in the host-cell membrane (West *et al.,* 2005; Kenjale *et al.*, 2005; Deane *et al.,* 2006; Matoo *et al.,* 2007; Dickenson *et al.,* 2011; Rahman & McFadden, 2011). Through this pore, the translocated proteins are able to enter the host cell cytoplasm where they subvert normal cell functions and induce major cytoskeletal re-arrangements to aid uptake of the bacterium. (Michinaga & Chichiro, 2006; Schroeder & Hilbi, 2008; Mounier *et al.*, 2009; Dickenson *et al.,* 2011).

Upon uptake, the bacterium is surrounded by the membrane vacuole of infected epithelial cells. After the vacuole membrane and escaping into the cytoplasm, *Shigella* are able multiply and spread to adjacent cells (Kaminski & Oaks, 2009; Marteyn *et al.*, 2010). Movement within the host cell is achieved by inducing local actin polymerization at one pole of the bacterium. (Mantis *et al.*, 1996; Faherty & Maurelli, 2009). This enables cell-to-cell spread without exposure to immune cells in the extracellular environment. Contact of *Shigella* with the membrane of the adjacent epithelial cell causes a protrusion to be formed and the bacterium is

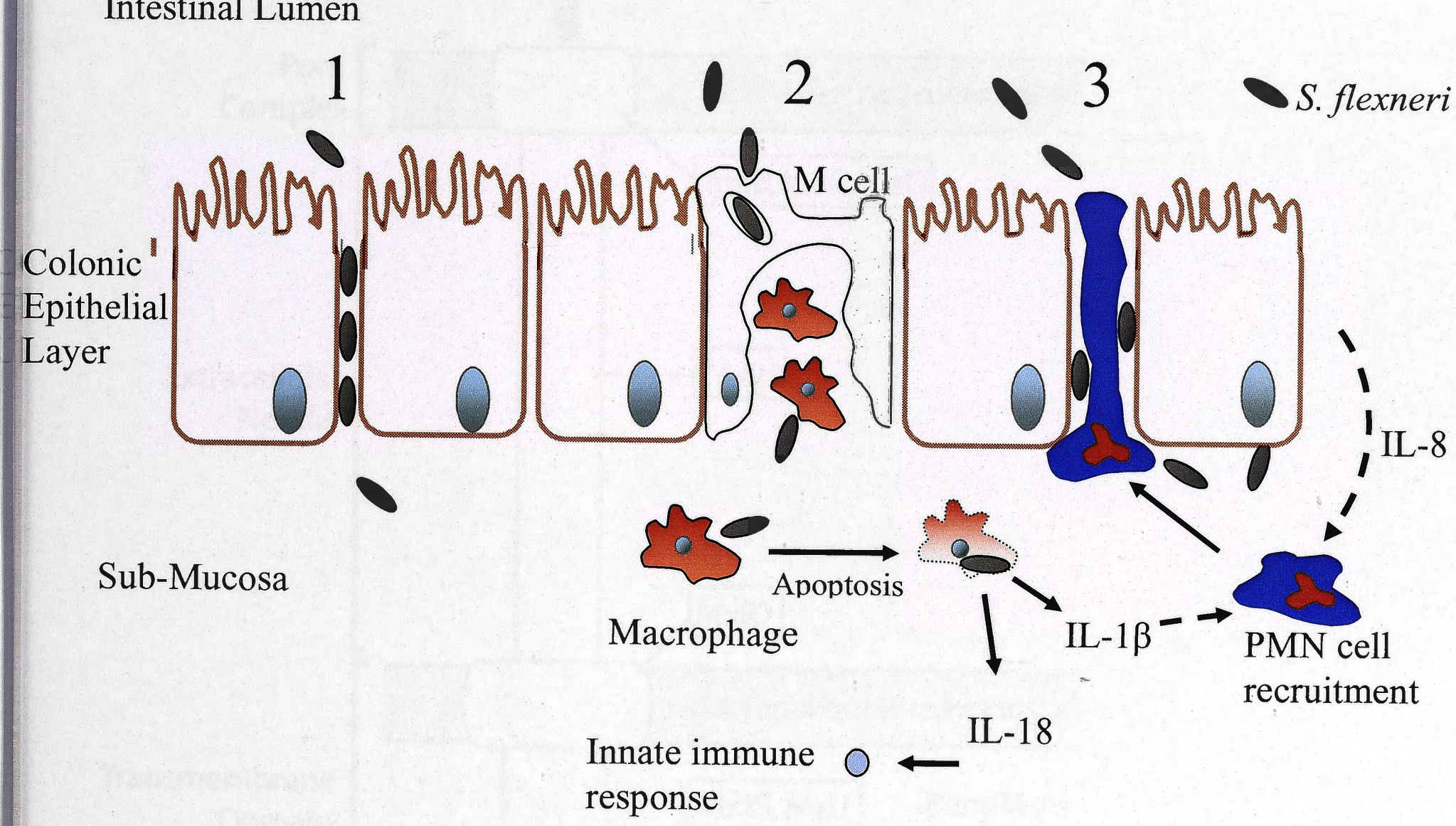


Figure 1.1: Invasion of the colonic epithelium by *Shigella*. Three mechanisms are employed by the bacteria in order to reach the basolateral side of the epithelial layer (sub-mucosa). 1) By manipulating the tight junction proteins; *Shigella* are able to mediate their passage into the sub-mucosa. 2) M cells facilitate entry by translocating bacteria in the intestinal lumen into the sub-mucosa. Proteins secreted by the bacteria cause the M-cell to phagocytose them. Bacteria are able to escape from the phagocytic vacuole and eventually from the M-cell. 3) The presence of *Shigella* in the intestinal lumen causes transmigration of polymorphonuclear cells (PMN) from the basolateral side to the intestinal lumen. The intercellular gaps created by the PMN cells are large enough for bacteria to pass through. Once in the sub-mucosa, the bacteria are confronted with resident macrophages. Apoptosis of the macrophages via the Caspase-1-dependent pathway leads to the release of cytokines IL-1β (chemotactic for PMN cells) and IL-18 (leads to proinflammatory singnalling and amplification of innate immune response). Continuous infection causes epithelial cells to undergo apoptosis resulting in ulceration and release of IL-8 (also causes PMN cell recruitment). Adapted from Schroeder and Hilbi (2008).

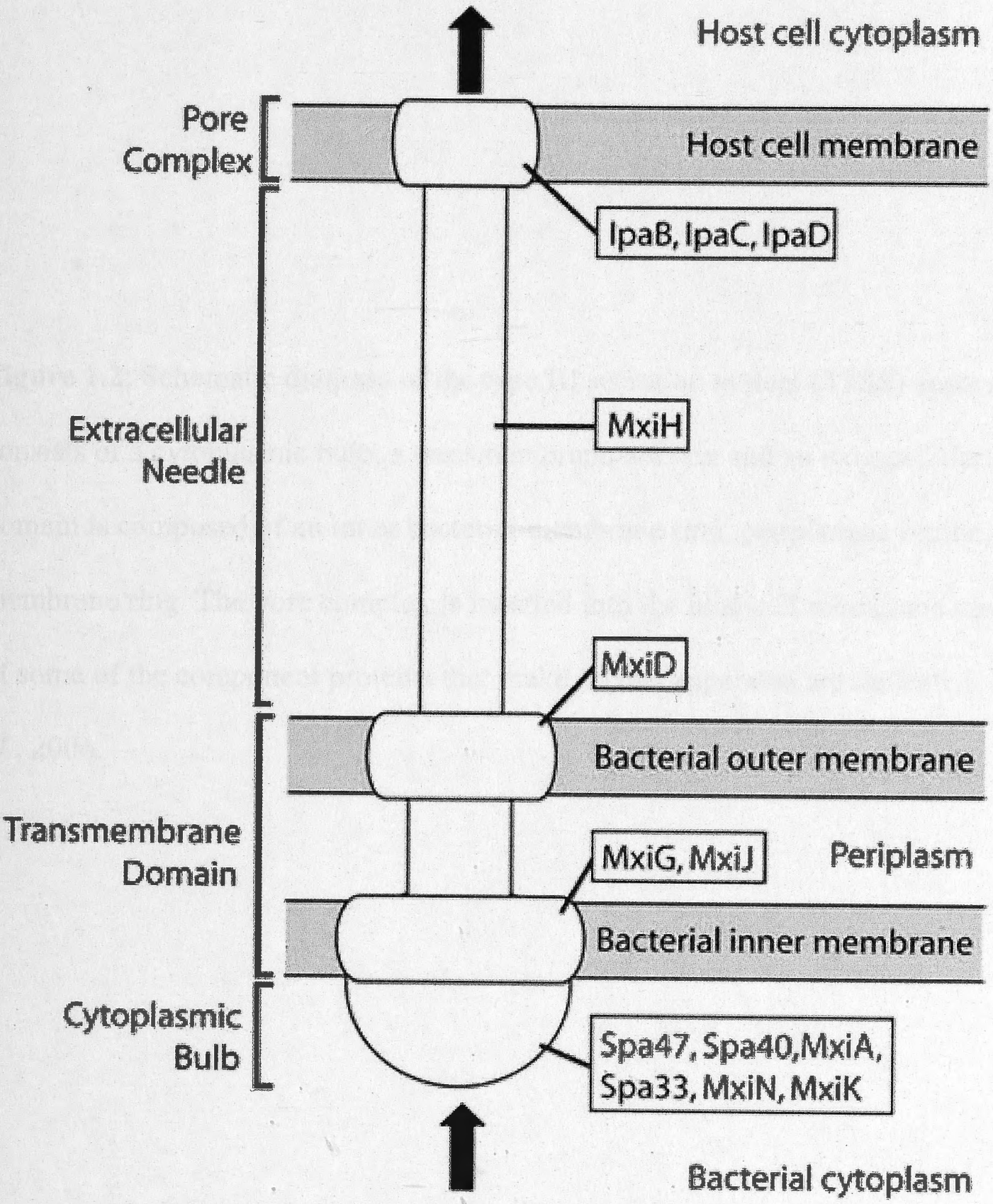

Host cell cytoplasm
Pore Complex
Host cell membrane
IpaB, IpaC, IpaD
Extracellular Needle
MxiH
MxiD
Bacterial outer membrane
Transmembrane Domain
MxiG, MxiJ
Periplasm
Bacterial inner membrane
Cytoplasmic Bulb
Spa47, Spa40, MxiA, Spa33, MxiN, MxiK
Bacterial cytoplasm

Figure 1.2: Schematic diagram of the type III secretion system (T3SS) apparatus. This secretion apparatus consists of a cytoplasmic bulb, a transmembrane domain and an extracellular needle. The transmembane domain is composed of an inner bacterial membrane ring, periplasmic region, and outer bacterial membrane ring. The pore complex is inserted into the host-cell membrane upon activation. The positions of some of the component proteins that make up this apparatus are indicated. Image taken from Deane *et al.*, 2006.

endocytosed by the neighbouring cell (Rathman *et al.*, 2000). The bacteria subsequently end up in a vacuole with two membranes, both of which are lysed by the proteins IpaB and IpaC (Espina *et al.*, 2006; Faherty & Maurelli, 2009). The major steps involved in *Shigella* invasion, replication and spread are summarised in Figure 1.3. It is thought that the host's inflammatory response is responsible for the killing of epithelial cells as there would be little advantage for *Shigella* in destroying an environment that provides protection from immune cells and favourable conditions for replication (Mantis *et al.*, 1996; Faherty & Maurelli, 2009).

1.2.1 Factors that determine Shigella *virulence*

The bacterial DNA of *Shigella* encodes a number of plasmid and chromosomal proteins such as the invasion plasmid antigens (Ipa), surface presentation antigens (Spa), membrane excretion proteins (Mxi), and virulence proteins (Vir). These proteins are encoded by genes that are scattered throughout the 221-kb virulence plasmid and are subsequently secreted through the T3SS for post-invasion virulence (Buchrieser *et al.*, 2000; Faherty & Maurelli, 2009) (Figure 1.4). A 31kb locus in this plasmid contains genes for entry (Sansonetti P, 2001; Schroeder & Hilbi, 2008; Faherty & Maurelli, 2009). This 31kb region consists of two main loci: the *mxi-spa* locus and *ipa* locus (Figure 1.5). There are about 20 genes associated with the *mxi-spa-ipa* locus encoding the T3SS. Temperature regulation of the genes on the virulence plasmid allows the needle complex to be synthesized and assembled at 37°C (Deane *et al.*, 2006; Matoo *et al.*, 2007; Faherty & Maurelli, 2009). The virulence protein VirF induces the expression of the VirB protein. The VirB protein then activates the *ipa, mxi,* and *spa* promoters leading to expression of the *spa* and *mxi* operons. This results in the synthesis and assembly of a protein complex called the Mxi-Spa translocon (Sansonetti P, 2001; Schroeder & Hilbi, 2008). At the basal state, the

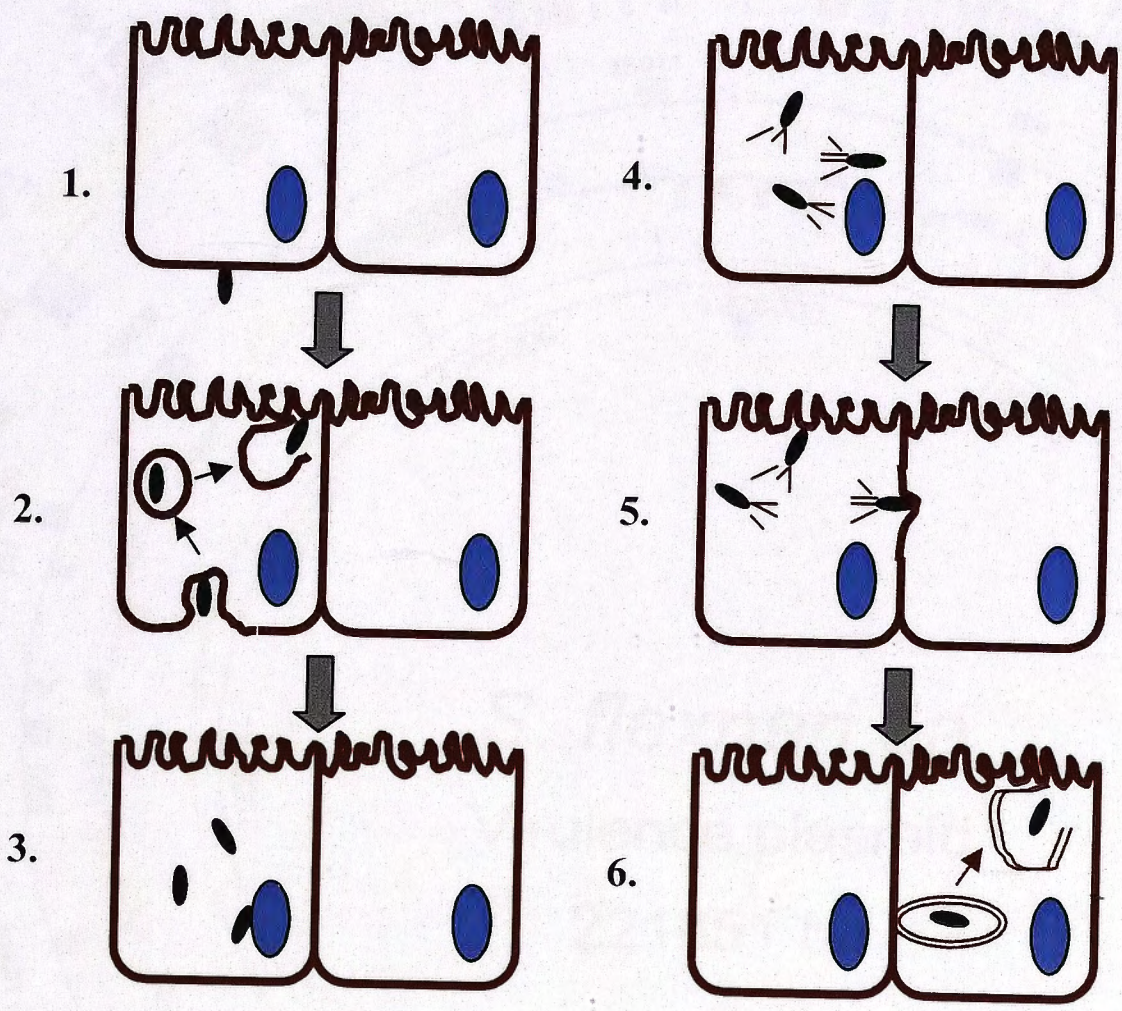


Figure 1.3: A summarized account of the major steps involved for *Shigella* invasion, replication and spread. 1) The bacterium makes contact with cell surface. 2) Invasion of the epithelial cell with the help of the TypeIII section system (T3SS). 3) Escape from the vacuole by lysis. 4) Replication of bacterium in cytosol. 5) Assembly of F-actin on one of the poles to enable intracellular movement. 6) Intercellular spread and lysis of double-membrane vacuole to allow the bacterium to escape. Adapted from Jennison & Verma, 2004.

Figure 1.4: Circular map of the large virulence plasmid (~221kb) of *Shigella flexneri* 5a. With respect to the outermost ring, all genes coloured in red are identical or essentially identical to known virulence-associated proteins. The 31kb mxi-spa-ipa locus is indicated with the help of the bold red bracket. Image taken from Venkatesan *et al.,* (2001)

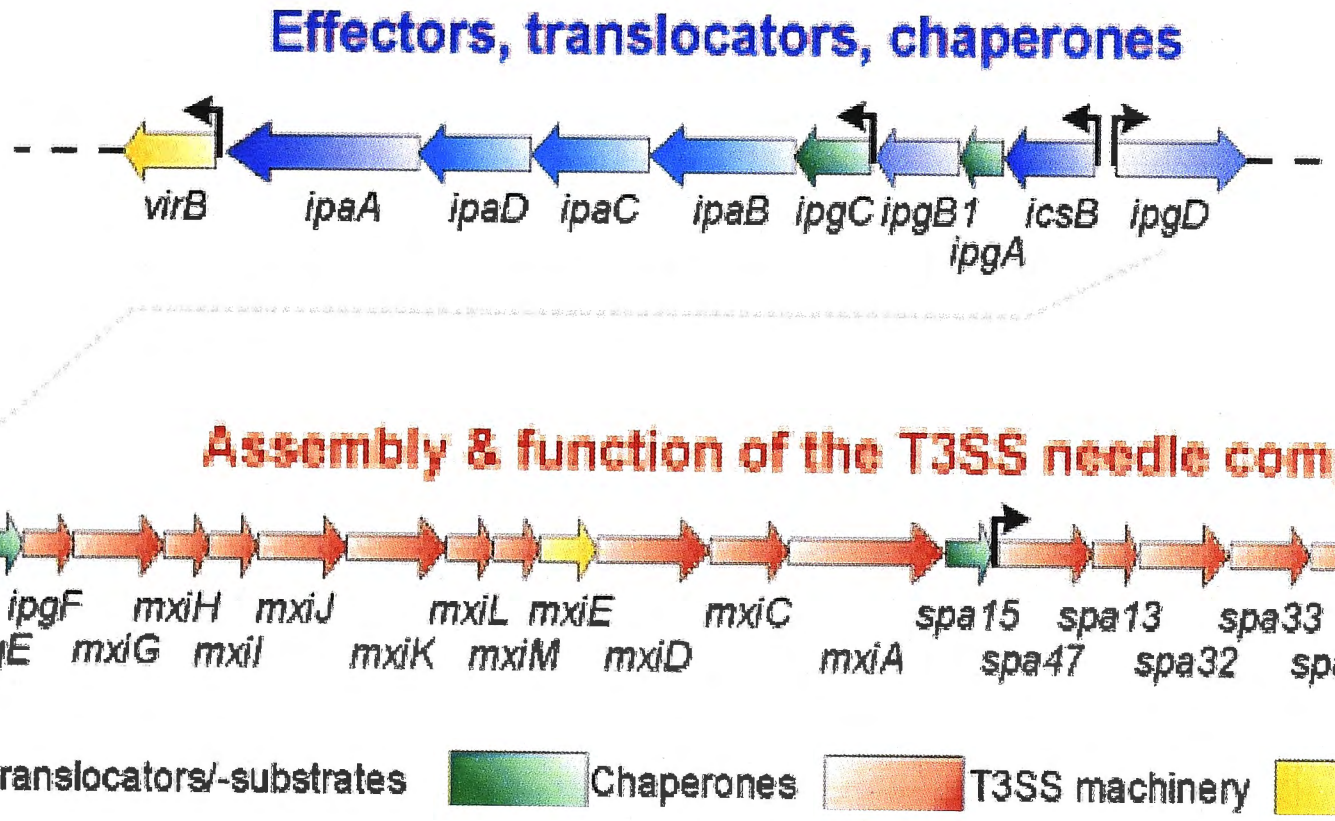


Figure 1.5: Map of the plasmid-located *Shigella mxi-spa-ipa* locus (31kb). This locus encodes the components needed for the assembly and function of a T3SS. This "pathogencity island" is required for entry into epithelial cells and killing of macrophages. Image taken from Schroeder & Hilbi (2008)

T3SS is inactive and presents at its tip two type III substrates, IpaB and IpaD, which form a so-called ''tip complex'' (Veenendaal *et al.,* 2007). The tip complex proteins regulate the activity of the T3SS and could participate in the recognition of host cell receptors, such as b1-integrins or the hyaluronan receptor CD44 (Parsot *et al.,* 1995; Watarai *et al.,* 1997; Lafont *et al.,* 2002; Veenendaal *et al.,* 2007). When the bacterium makes contact with the epithelial cell membrane, the translocon becomes activated and secretes the pre-synthesized Ipa proteins. IpaB, IpaC and IpaD associate to form a complex which interacts with the host epithelial cell membrane to induce a cascade of cellular signals which will lead to the internalization of the bacterium via an endosome. The Ipa proteins are also required for escape from the endosome (Nhieu *et al.,* 1997; Jennison & Verma, 2004; Gala´ n & Wolf-Watz, 2006; Ogawa et al., 2008). Other genes within this locus include *spa40*, which encodes for one of the inner membrane components, Spa40. This protein belongs to a family of proteins that regulate the switching of substrate specificity of the T3SS (Deane *et al.,* 2008). Spa15 (encoded by *spa15*) has been described as a T3SS chaperone based on the requirement of Spa15 for secretion of several proteins (Parsot *et al.,* 2005). However, recent studies have revealed that Spa15 contributes to the intracellular survival of *Shigella* by blocking apoptosis in the infected host cell (Faherty & Maurelli, 2009).

Shigella pathogenicity islands (SHI) have also been identified in the chromosome (Ingersoll *et al.,* 2002; Schmidt & Hensel, 2004). The presence and genomic localization of SHI varies between *Shigella* strains and may contribute to the variation of virulence phenotypes (Peng *et al.,* 2006). Although these islands contain a host of genes, specific virulence functions have only been identified for some of these genes. Table 1.1 lists these various SHI virulence factors that have been identified to date. The acquisition of SHI-O was a very important event in

Table 1.1: Virulence factors encoded by chromosomal *Shigella* PAIs (Schroeder & Hilbi, 2008)

PAI	Gene(s)	Biochemical activity	Virulence function(s)
SHI-1	*sigA*	Immunoglobulin A–like cytopathic protease	Intestinal fluid accumulation, cytopathic toxin
	pic	Serine protease, mucinase	Mucus permeabilization, serum resistance, hemagglutination
	set1A, *set1B*	Enterotoxin ShET1	Intestinal fluid accumulation, development of watery diarrhea
SHI-2	*iucA* to *iucD*, *iutA*	Siderophore (aerobactin) production and transport	Iron acquisition
	shiD	Unknown	Colicin I and colicin V immunity
	shiA	Unknown	Downregulation of inflammation by suppression of T-cell signaling
SHI-3	*iucA* to *iucD*, *iutA*	Aerobactin production and transport	Iron acquisition (found only in *S. boydii*)
SHI-O	*gtrA*, *gtrB*, *gtrV*	Modification of O antigens, serotype conversion	Evasion of host immune response
SRL	*fecA* to *fecE*, *fecI*, *fecR*	Ferric dicitrate uptake	Iron acquisition
	tetA to *tetD*, *tetR*	Efflux system	Tetracycline resistance
	cat	Acetyltransferase	Chloramphenicol resistance
	oxa–1	β-Lactamase	Ampicillin resistance
	aadA1	Adenylyltransferase	Streptomycin resistance

the evolution and diversification of *Shigella* (Schroeder & Hilbi, 2008). This island contains the genes that are responsible for carrying out O-antigen modification and as a result, accounts for the wide variety of *Shigella* serotypes (Nie *et al.,* 2006). A PAI, designated the *Shigella* resistance locus (SRL), is the most recent acquisition of genes and has contributed to the organism's resistance to antibiotics such as ampicillin, chloramphenicol, tetracycline and streptomycin (Turner *et al.,* 2001; Turner *et al.,* 2003).

1.3 Prevention and Treatment of shigellosis

The best means of preventing shigellosis is the usage of clean water and sanitary facilities. However in developing countries the use of such ideals is far from reaching. Thus other means for preventing and treating the disease are required.

Treatment of acute or self-limiting shigellosis involves oral re-hydration to restore lost levels of water in the stool. In more severe cases, antibiotic therapy is required. Using antibiotics in the treatment of shigellosis can shorten the period of faecal excretion of *Shigella,* thereby reducing the possibility of secondary spread of the infection and limiting the course of illness. Antibiotic therapy often involves the use of ampicillin, trimethoprim–sulfamethoxazole and tetracycline (Ashkenazi *et al.*, 2003; Christopher *et al.,* 2010). However, *Shigella* strains are becoming more resistant to these antibiotics through the uptake of plasmids containing antibiotic resistant genes, thus requiring the development of new antibiotics (Barman *et al.*, 2010). This, coupled with the high cost of antibiotic treatments, has led to WHO giving priority to the development of an effective vaccine against shigellosis (Kotloff *et al.,* 1999)

After an initial *Shigella* infection, a 70-80% protection against reinfection from the same serotype is induced (Kotloff *et al.*, 1999). Therefore, vaccination is a viable option for the control of shigellosis. Although the mechanisms of protection against *Shigella* have not been elucidated, mucosal immune defenses are sought, as they block pathogens in the intestinal lumen, preventing their attachment and infection of epithelial cells. Therefore, the optimal *Shigella* vaccine must be able to activate long-lasting mucosal immunity, be cheap to produce, cause minimal side effects and be simple to administer (Lindberg & Pal, 1993; Salam, 1998). Most of the vaccines currently available have low immunogenic responses or are not well tolerated (Kaminski & Oaks, 2009). A common problem faced during the development of vaccines against shigellosis is that immunity is serotype specific. As there are multiple serotypes present in a single region, a vaccine produced against a single serotype would not be beneficial. Thus, research into development of a multi-valent vaccine which favours long-lived, broad immunity is needed (Kaminski & Oaks, 2009).

Many different vaccine strategies such as subunit vaccines, killed oral vaccines, non-invasive live vaccines, invasive live vaccines, and *Shigella-E. coli* hybrid vaccines have been adopted and tested in animals and humans (Kweon M. N., 2008; Kaminski & Oaks, 2009; Martinez-Becerra *et al.,* 2012). However, the achievement of a balance between immunogenicity and safety has been problematic. Oral immunization may achieve intestinal immunity but due to delivery restrictions, it is often difficult to consistently accomplish especially for inactivated or subunit vaccines (Kaminski & Oaks 2009; Tribble *et al.,* 2010). An alternative route of delivery, touted to trigger the mucosal immune response is through intranasal immunization (Tribble *et al.,* 2010; Riddle *et al.,* 2011).

1.3.1 Promising vaccine candidates

In recent years, there have been some promising vaccine candidates. Protein antigens, such as the Ipa protein effectors of the type three-secretion system, remain attractive vaccine candidates due to their active role in pathogen-directed cellular invasion, as well as the highly conserved sequences among these essential virulence factors (Kweon M. N., 2008; Kaminski & Oaks, 2009; Tribble *et al.,* 2010; Riddle *et al.,* 2011). By exploiting the capacity of the *Shigella* Ipa proteins to interact with the host cell receptors and induce an endocytic event, researchers at the Walter Reed Army Institute of Research (Washington DC, USA) have developed an invasin complex (Invaplex 50) vaccine (Tribble *et al.,* 2010; Riddle *et al.,* 2011; Martinez-Becerra *et al.,* 2012). The Invaplex product, isolated from wild-type *S. flexneri* 2a, contains a native, biologically active complex of LPS, IpaB, IpaC and IpaD that targets M cells, resulting in the uptake and delivery of key *Shigella* antigens to underlying antigen presenting cells (APCs) (Kaminski *et al.,* 2009; Tribble *et al.,* 2010; Riddle *et al.,* 2011). The presence of LPS and the conserved Ipa proteins stimulates a more comprehensive immune response similar to that witnessed after natural infection compared to vaccines that deliver LPS alone (Kaminski *et al.,* 2009; Riddle *et al.,* 2011; Martinez-Becerra *et al.,* 2012). The *Shigella* Invaplex 50 vaccine, currently under clinical investigation, was seen to stimulate robust intestinal and pulmonary responses upon intranasal immunization. Encouragingly, an initial phase 1 dose-escalation study evaluating the safety and immunogenicity of the *S. flexneri* Invaplex 50 vaccine showed that it was well tolerated and resulted in antigen-specific humoral and mucosal immune responses in adults (Riddle *et al.,* 2011).

In another study conducted by Martinez-Becerra *et al.* (2012), they examined the immunogenicity and protective efficacy of IpaB and IpaD, alone or combined, co-administered with a double mutant heat-labile toxin (dmLT) from *E. coli* as a mucosal adjuvant in a mouse model of intranasal immunization and pulmonary challenge. The authors demonstrated that both proteins were able to induce high levels of serum and mucosal antibodies, along with elevated frequencies of antibody secreting cells in mucosal and systemic tissues. It was observed that very high levels of protection (80 to 100%) were achieved against a homologous and a heterologous strain using a lethal pulmonary challenge model. Unlike the Invaplex 50, this vaccine does not contain LPS. It also only contains known quantities of highly purified IpaB, along with IpaD (Martinez-Becerra *et al.*, 2012). An advantage to this vaccine candidate is the fact that a protein-based, LPS-free vaccine would be safer and amenable for use in infants, young children and immune-compromised individuals (Martinez-Becerra *et al.*, 2012). As such, the study concluded that subunit-based IpaB/D vaccine could be an effective broadly protective vaccine for the prevention of shigellosis (Martinez-Becerra *et al.*, 2012)

1.4 Evolution of *Shigella*

Over the last 15 years, studies employing a variety of technical approaches have shown that *Shigella* belong to the species *E. coli*, thus debunking the long held dogma that *Shigella* formed a separate genus (Pupo *et al.,* 1997; Fukushima *et al.,* 2002). In the past decade, new sequence information has revealed that the sequence divergence between *S. flexneri* and *E. coli* K-12 is about 1.5% compared to 15% diversity between *E. coli* and its close relative, pathogenic *Salmonella enterica* (Lan & Reeves, 2002; Lan *et al.,* 2004). Using eight housekeeping gene sequences from the *Shigella* chromosomes, Pupo *et al.* (2000) proposed that the current *Shigella*

clones were independently evolved from multiple *E. coli* ancestors 35,000– 270,000 years ago. Subsequent genome sequencing projects have strongly supported the multiorigin and convergent evolution hypothesis (Jin *et al.,* 2002; Wei *et al.,* 2003; Yang *et al.,* 2005). Bacterial evolution is a dynamic process whereby horizontal gene transfer between microorganisms of distant relation allows them to acquire genes that lead to the gain of advantageous phenotypic traits (Ochman *et al.,* 2000). The transfer and chromosomal incorporation of pathogenicity islands (PAI) are important steps for the rapid evolution of pathogenic *Shigella* bacteria from their non-pathogenic *E. coli* progenitors. Through the acquisition of a large virulence plasmid and chromosomal pathogenicity islands mentioned in the above section, *Shigella* spp. have gained genes that have enabled them to invade the epithelial cells of the host and to multiply and thrive intracellularly. Both the *Shigella* chromosome and the virulence plasmid contain numerous insertion sequences (IS) and markers of genomic rearrangements (Buchreiser *et al.,* 2000; Jin *et al.,* 2002). The distribution of IS types and the phylogenetic grouping of chromosome- and plasmid-encoded genes indicate that the *Shigella* chromosome and the virulence plasmid coevolved (Escobar-Paramo *et al.,* 2003; Yang *et al.,* 2007). As such, these findings suggest that the acquisition and transfer of plasmids between the strains were early steps in the diversification from *E. coli* ancestors.

While gaining genes that increase virulence and assist in the bacteria's intracellular survival has contributed significantly to the evolution of *Shigella*, it is also shaped by loss of gene function due to deletion or gene inactivation. Compared to *E. coli* K-12, an average of 726 genes are missing, and over 200 pseudogenes are found within *Shigella*, indicating that genes within *Shigella* have been deleted, reduced or inactivated (Yang *et al.,* 2005; Peng *et al.,* 2006).

Genome reduction due to the loss of functional genes is driven mainly by the bacteria adapting to its pathogenic intracellular lifestyle (Ochman *et al.,* 2000; Maurelli A. T., 2007). Examples of genes that have been inactivated include *ompT,* which encodes for the outer membrane protease OmpT. Studies have shown that it interfered with the polar localisation of the actin nucleator protein IcsA (VirG) which in turn, inhibited the spread and virulence of *Shigella* (Peng *et al.,* 2006). It is present in various non-pathogenic *E. coli* strains but deleted in all *Shigella* strains (Peng *et al.,* 2006). Genome analyses also confirmed that *cadA*, encodes a lysine decarboxylase, though present in ancestral *E. coli* strains, was deleted in most of the *Shigella* due the protein's indirect inhibitory effect on the activity of *Shigella* enterotoxins (Day *et al.,* 2001; Peng *et al.,* 2005). In addition to loss of genes, a very prominent trait that lost by *Shigella* is the ability to synthesise functional flagella. Tominaga *et al.* (1994) identified flagellar genes in *Shigella* that were expressed in E. coli K-12. Evidence has also been shown of flagella and motility in *Shigella* species under modified laboratory conditions (Giron, J. A, 1995). Flagella and other surface structures like fimbriae, which are also degenerated in *Shigella*, have been found to be potent activators of the host immune system (Sakellaris *et al.*, 2000; Ramos *et al.,* 2004).). Therefore, the loss of these surface structures might may minimize the exposure to the host defence system and allow the bacterium to invade the epithelial cells more efficiently. Figure 1.6 documents the events that are considered to be most important for the evolution of pathogenic *Shigella* from non-pathogenic *E. coli*.

A peek into the evolution of *Shigella* could be provided by studying enteroinvasive *Escherichia coli* (EIEC). EIEC is a pathogenic form of *E. coli* that can cause dysentery. However, unlike non-pathogenic *E. coli*, it is difficult to make a distinction between EIEC and

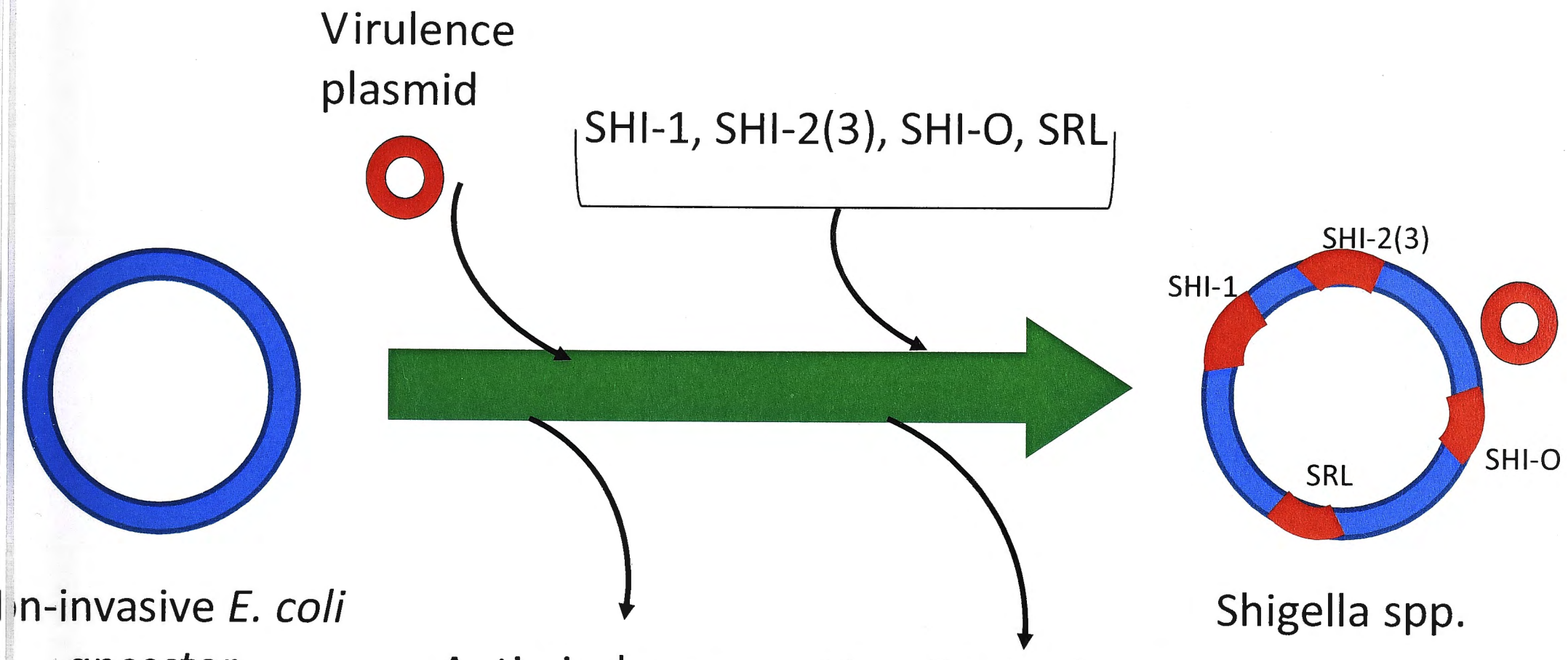


Figure 1.6: Genetic events that contributed to the evolution of *Shigella* spp. from their non-invasive *E. coli* ancestor. This was achieved over time, through the acquisition of the virulence plasmid and various chromosomal pathogenicity islands. In that time, genetic loci that impeded virulence or were not functional intracellularly were lost. SRL-*Shigella* resistance locus. Adapted from Schroeder & Hilbi, (2008).

Shigella as both EIEC and *Shigella* bear almost identical phenotypic likeness from spending much of their lifetime within eukaryotic cells (Lan *et al.,* 2004). Therefore, both organisms have a different nutrient supply from most *E. coli* strains and harbor a common 221-kb virulence plasmid (Lan *et al.,* 2004). However, EIEC and *Shigella* lack several catabolic pathways widely present in *E. coli have* (Lan *et al.,* 2004). Mucate and acetate are an example of two biochemical properties that differentiate *E. coli* and *Shigella*. EIEC may be positive for one or both of the properties, while, with rare exceptions, *Shigella* strains are negative for both and more than 90% of typical *E. coli* strains are positive for both (Doyle & Padhye 1989; Bopp *et al.,* 2003; Lan *et al,* 2004). It is clear that EIEC and *Shigella* form a pathovar of *E. coli*. As such, Lan *et al.,* (2004) have speculated that the reason EIEC retained some *E. coli* properties that have been lost in multiple lineages of *Shigella* could be due the EIEC strains being an intermediate stage and as such, may be a precursor of "full-blown" *Shigella* strains.

1.5 Serotype diversity of *Shigella flexneri*

To date, there are 15 known *S. flexneri* serotypes, which differ in their virulence, prevalence, and distribution. The *S. flexneri* serotypes differ in their O-antigens. These are components of lipopolysaccharide that extend from the bacterial surface.

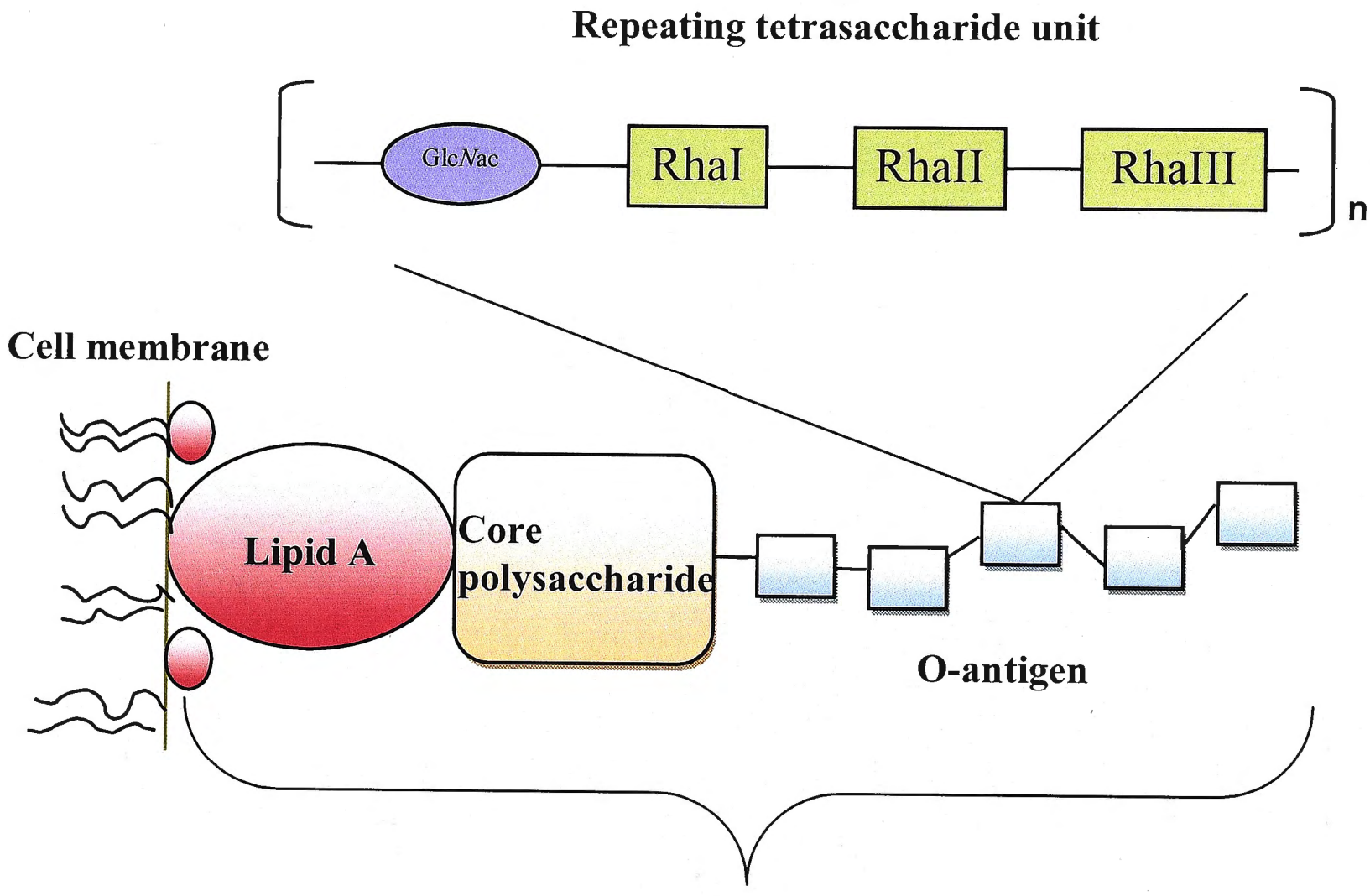
1.5.1 Lipopolysaccharide (LPS) structure

The bacterial LPS is a major constituent of outer-membrane structure that occupies most of the cell wall of Gram-negative bacteria and is vital to their virulence (Guan *et al.*, 1999; Nagy & Pal, 2008). It accounts for approximately 10-15% of the total molecules in the outer membrane occupying approximately 75% of the bacterial surface area (Lerouge &

Vanderleyden, 2002). Most Gram-negative bacteria express a common LPS backbone, which consists of; a lipid A, core polysaccharide and O-antigen (Allison & Verma, 2000) (Figure 1.7). The lipid A, which is anchored in the outer membrane, is highly conserved in members of the *Enterobacteriaceae* family and is responsible for maintaining the integrity of the outer membrane LPS structure (Lerouge & Vanderleyden, 2002; Acheson & Luccioli, 2004). The core polysaccharide links lipid A to the O-antigen and is composed of a short chain of sugars. The core polysaccharide has slightly more variation between members of the *Enterobacteriaceae* family, but it is still limited in structural diversity (Lerouge & Vanderleyden, 2002; Nagy & Pal, 2008). The O-antigen on the other hand (outermost structure of LPS) is highly variable and gives rise to different *S. flexneri* serotypes. It extends outward from the core polysaccharide and can consist of up to 40 repeating subunits comprising of three to five sugars. Serotype diversity is brought about by the differences in the sugar residues, their relative position and linkage.

1.5.2 Properties and synthesis of the Shigella flexneri *O-antigen*

The O-antigen is highly immunogenic and can elicit a strong immune response due to the fact that it is the only part of the bacterial LPS that is exposed to the environment (Opal, 2007; Carter *et al.*, 2007; Phalipon & Sansonetti, 2007; Nagy & Pal, 2008). The length of the O-antigen varies and has been found to change depending on the growth phase of the bacteria (Carter *et al.*, 2007). Studies have shown that the length of the LPS is fundamental to the adherence and internalisation of *S. flexneri* into polarized epithelial cells (Carter *et al.*, 2007). The O-antigen has also been found to be critical to the survival of Gram-negative bacteria as it prevents penetration from the complement membrane attack complex (MAC) of the innate immune response, thus preventing the clearance of the bacteria by the complement system of the innate immune



Lipopolysaccharide

Figure 1.7: The basic bacterial lipopolysaccharide (LPS) structure consisting of a lipid A, core-polysaccharide and O-antigen. The lipid A anchors the core-polysaccharide and O-antigen to the outer cellular membrane. The O-antigen is different for each *S. flexneri* serotype. The basic Y serotype O-antigen is shown here (Glc*N*ac, N-acetylglucosamine; Rha, rhamnose).

response (Erridge *et al.*, 2002). It has been proposed that modifications to the O-antigen via glucosylaton would induce it to change from a linear to helical conformation with the glucosyl residue exposed on the exterior of the helix. This would dramatically shorten the O-antigen, halving the distance it extends beyond the outer membrane (West *et al.,* 2005) and may affect the accessibility of the type III secretion system that is required for bacterial invasion (West *et al.,* 2005). Though the serotypes differ in their virulence, it is not yet known to what extent the O-antigen structure contributes to this.

There are three distinct processes for the synthesis and translocation of O-antigens, namely, the Wzx/Wzy dependent process, the ATP-binding cassette (ABC) transporter dependent process and the synthase dependent process (Samuel & Reeves, 2003; Liu *et al.,* 2008). Most *E. coli* and all the *Shigella* forms use the Wzx/Wzy process (Figure 1.8). The initial step in the Wzy-dependant pathway model involves synthesis of the repeating polysaccharide units on the lipid carrier molecule undecaprenyl phosphate (UndP) on the cytoplasmic leaflet of the inner membrane (IM). Following this, the O-antigen flippase, Wzx, translocates the UndP-linked sugar repeats from the cytoplasmic leaflet to the periplasmic leaflet (Liu *et al.,* 1996; Burrows & Lam, 1999). Here, the repeating units are sequentially added to the reducing terminus of a growing glycan chain by the O-antigen polymerase Wzy (de Kievit *et al.,* 1995; Woodward *et al.,* 2010). The length of the growing chain is governed by the chain length regulator Wzz (Woodward *et al.,* 2010). The assembly process is then completed when the extended O-antigen chain is ligated to the lipid A-core polysaccharide complex by the O-antigen ligase WaaL (Abeyrathne & Lam, 2007) which the O unit is synthesized by transfer of a sugar phosphate and then sequential transfer of other sugars from the respective sugar nucleotides to the carrier lipid.

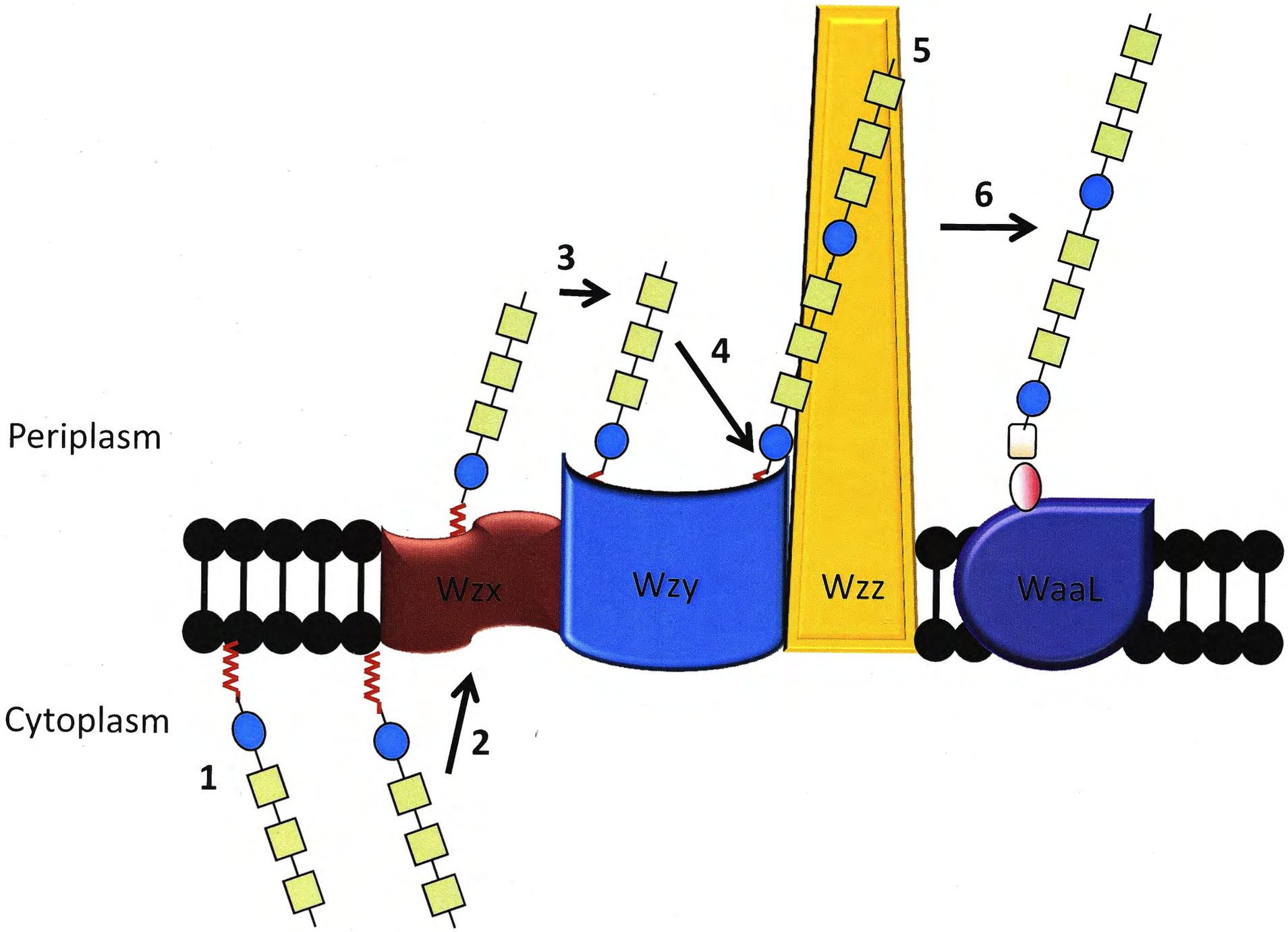

Periplasm
Cytoplasm
1
2
3
4
5
6
Wzx
Wzy
Wzz
WaaL

Figure 1.8: Wzy-dependent assembly pathway. 1) Synthsis of O-antigen polysaccharide repeat unit on UndP by glycosyltransferases. 2) The flippase, Wzx, translocates the UndP-linked O-antigen from the innner leaflet to the outer leaflet. 3) UndP-linked O-antigen binds to recruitment arm of O-antigen polymerase Wzy. 4) Single UndP-linked O-antigen repeat transferred from recruitment arm of to the base of the retention arm of Wzy. This leads to the polymerization of the glycan chain at the reducing terminus. 5) Wzz regulates the length of the O-antigen chain. 6) Polymerised O-antigen is ligated to the lipd A-core polysaccharide complex. This gives rise to a heteromeric LPS. Adapted from Cuthbertson *et al.,* (2010)

1.5.3 O-antigen structure of the known Shigella flexneri *serotypes*

The 15 known serotypes of *S. flexneri* differ in the nature of their O-antigens. The O-antigen backbone consists of repeating units of the tetrasaccharide *N*-acetylglucosamine-rhamnose-rhamnose-rhamnose (Simmons and Romanowska, 1987). This basic backbone is present for all serotypes except serotype 6. Bacteria with this basic O-antigen repeat unit have been typed as serotype Y. The addition of glucosyl and/or O-acetyl groups to different sugars in the tetrasaccharide unit by one of several linkages gives rise to the other serotypes (Figure 1.9). Glucosylation can occur in any one of the four residues that are present in the tetrasaccharide unit. However, O-acetylation occurs only on the rhamnose (Rha) III residue, resulting in the presence of the group 6 epitope (Jennison *et al.,* 2006).

Serotypes 1a, 2a, 4a, 5a and X differ from serotype Y by the addition of a single glucosyl group. Serotypes 2b and 5b have two glucosyl groups added on different acceptor sugar residues. Serotype 3b has a single O-acetyl group added. Serotypes 1b, 3a and 4b have both an O-acetyl and a glucosyl group added. Serotype 1c is unique as it has two glucosyl groups added sequentially with a direct glucosyl to glucosyl linkage (Ramiscal *et al.*, 2010). Each serotype has one type-specific, and one or more group-specific, antigenic determinants (Jennison *et al.,* 2006). The type antigen denotes a sugar residue with a specific linkage to the basic O-antigen chain (serotypes Y and X do not have type antigens), while the group antigen refers to an additional sugar or a particular immunogenic chemical conformation. This scheme for classifying serological specificity was invented before the chemical structure of the O-antigen was determined, and was based on the use of rabbit antisera (Kenne *et al.*, 1978).

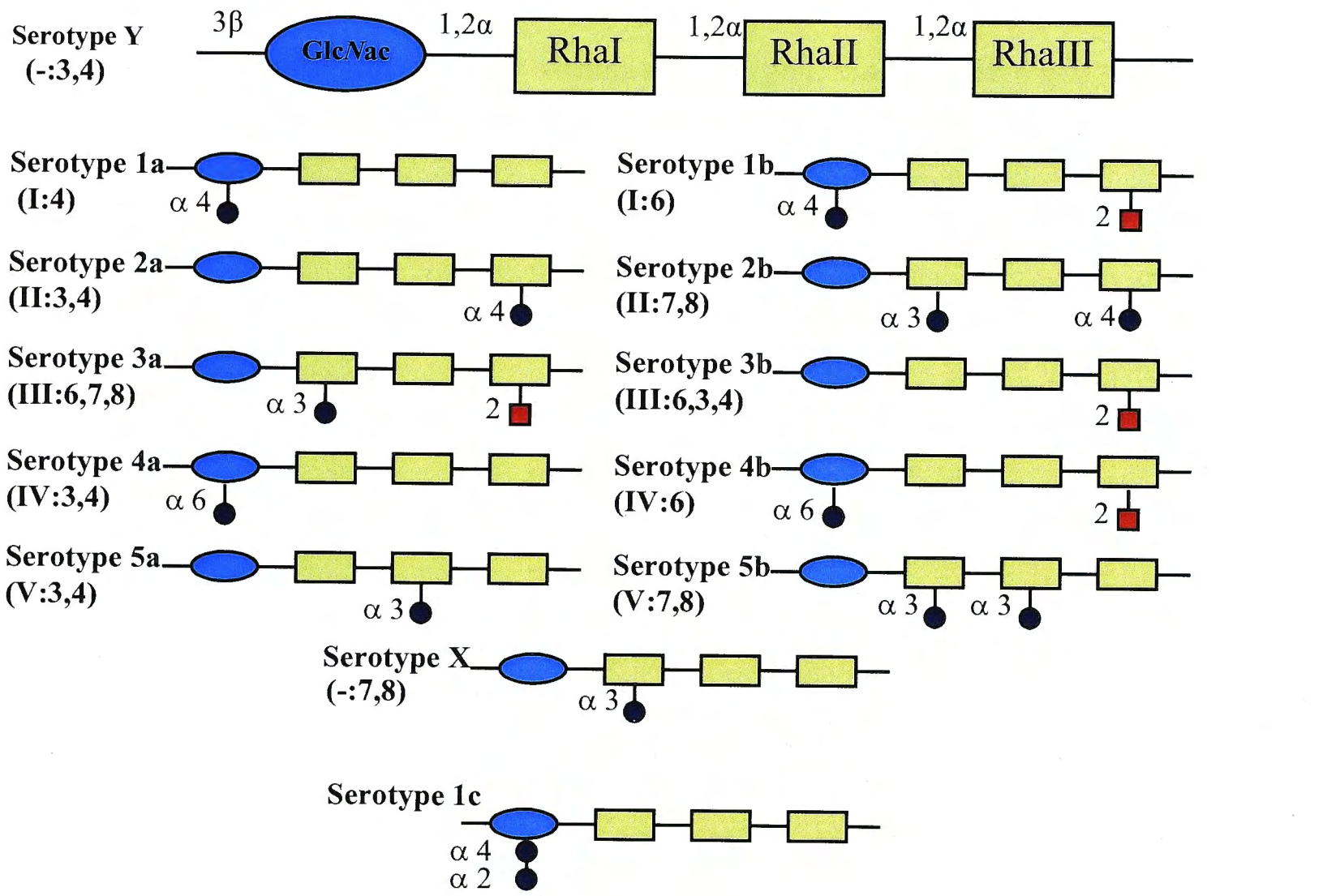


Figure 1.9: The different serotypes of *S. flexneri* and their respective O-antigens. The basic O-antigen (Serotype Y) is a tetrasaccharide consisting of repeating N-acetylglucosamine-rhamnose (I)-rhamnose (II)-rhamnose (III) units. The large variety of serotypes is dependent on the attachment of glucosyl groups (glucosylation) and/or O-acetyl groups (acetylation) in a site and linkage specific manner to different sugar molecules in the O-antigen. Serotype 4a has a glucosyl group attached to the N-acetylglucosamine via an α1-6 linkage. Adapted from Allison & Verma (2000)

1.5.4 Modification of the O-antigen

Serotype conversion in *S. flexneri* is mediated by O-antigen modification factors encoded by temperate bacteriophages. It is not unusual for bacteriophages that utilise O-antigens as receptors for adsorption and infection to contain O-antigen modification genes (Makela and Stocker, 1969). These bacteriophages may possess glucosyltransferase genes, acetyltransferase genes, or novel O-antigen polymerase systems to modify the O-antigen of the bacterial host (Wright, 1971; Sasaki *et al.*, 1974; Verma *et al.*, 1991). The bacteriophages most likely modify the O-antigen of their bacterial host to exclude the entry of homologous bacteriophages and/or to contribute to their host bacterium's evasion of the immune response (Guan *et al.*, 1999).

1.5.4.1 Serotype converting bacteriophages

Bacteriophages are viruses that infect bacteria, and can be important determinants of bacterial pathogenesis and evolution. In the 125 completed bacterial genome sequences, at least 300 intact or defective prophages (integrated bacteriophage genomes) have been identified (Casjens *et al.*, 2004). Six different serotype-converting phages or prophages, SfI, SfII, Sf6, SfIV, SfV and SfX, have been identified and characterized which can convert serotype Y to serotype 1a, 2a, 3b, 4a, 5a and X, respectively (Clark *et al.,* 1991; Mavris *et al.,* 1997; Guan *et al.,* 1999; Allison *et al.,* 2002; Casjens *et al.,* 2004; Sun *et al.,* 2011). Except for Sf6 which carries a single gene, *oac*, for acetylation of the O-antigen, the other phages carry three genes, *gtrA*, *gtrB*, and $gtr_{(type)}$ for O-antigen modification (Verma *et al.,* 1991). The morphologies of these bacteriophages are diverse, yet their integration, excision and glucosylation modules are conserved (Allison and Verma, 2000). Sf6 does not have a glucosylation module but encodes O-acetyltransferase, and modifies the O-antigen of its host by the addition of the O-acetyl groups

(Verma *et al.*, 1991). The glucosylation modules in serotypes 1a and 4a are found in non-inducible bacteriophage remnants (Adhikari *et al.*, 1999; Adams *et al.*, 2001).

The bacteriophages responsible for serotype conversion in *S. flexneri* establish a stable relationship (lysogeny) with the host bacterium, and are therefore temperate bacteriophages. Site-specific recombination between the bacteriophage and the *S. flexneri* genome occurs at short homologous DNA sequences in the bacteriophage and bacterial genomes called attachment sites (*attP* and *attB*, respectively). The process is in accordance with the Campbell model first described for bacteriophage λ (Campbell, 1986).

1.5.4.2 O-antigen glucosylation

Most of the *S. flexneri* serotypes arise from the addition of a glucosyl residue to a specific sugar of the O-antigen repeat unit via one of several linkages. The three genes involved in O-antigen glucosylation are *gtrA*, *gtrB*, and $gtr_{(type)}$. These genes are located downstream of the *attP* site in the bacteriophage genome (Allison and Verma, 2000). While *gtrA* and *gtrB* are highly conserved and interchangeable among serotypes (Huan *et al.*, 1997; Mavris *et al.*, 1997; Guan *et al.*, 1999), $gtr_{(type)}$ is serotype-specific and is unique to each serotype. The organisation of the glucosylation gene clusters that have been identified up till now is illustrated in Figure 1.10.

gtrA is highly conserved among the serotypes and encodes a 120-amino-acid protein with four transmembrane helices (Korres et al., 2005). Guan et al. (1999) proposed that GtrA is involved in flipping lipid-linked glucose across the plasma membrane for it to be available to the Gtr(type) for the attachment of the glucose to the correct rhamnose unit conferring serotype

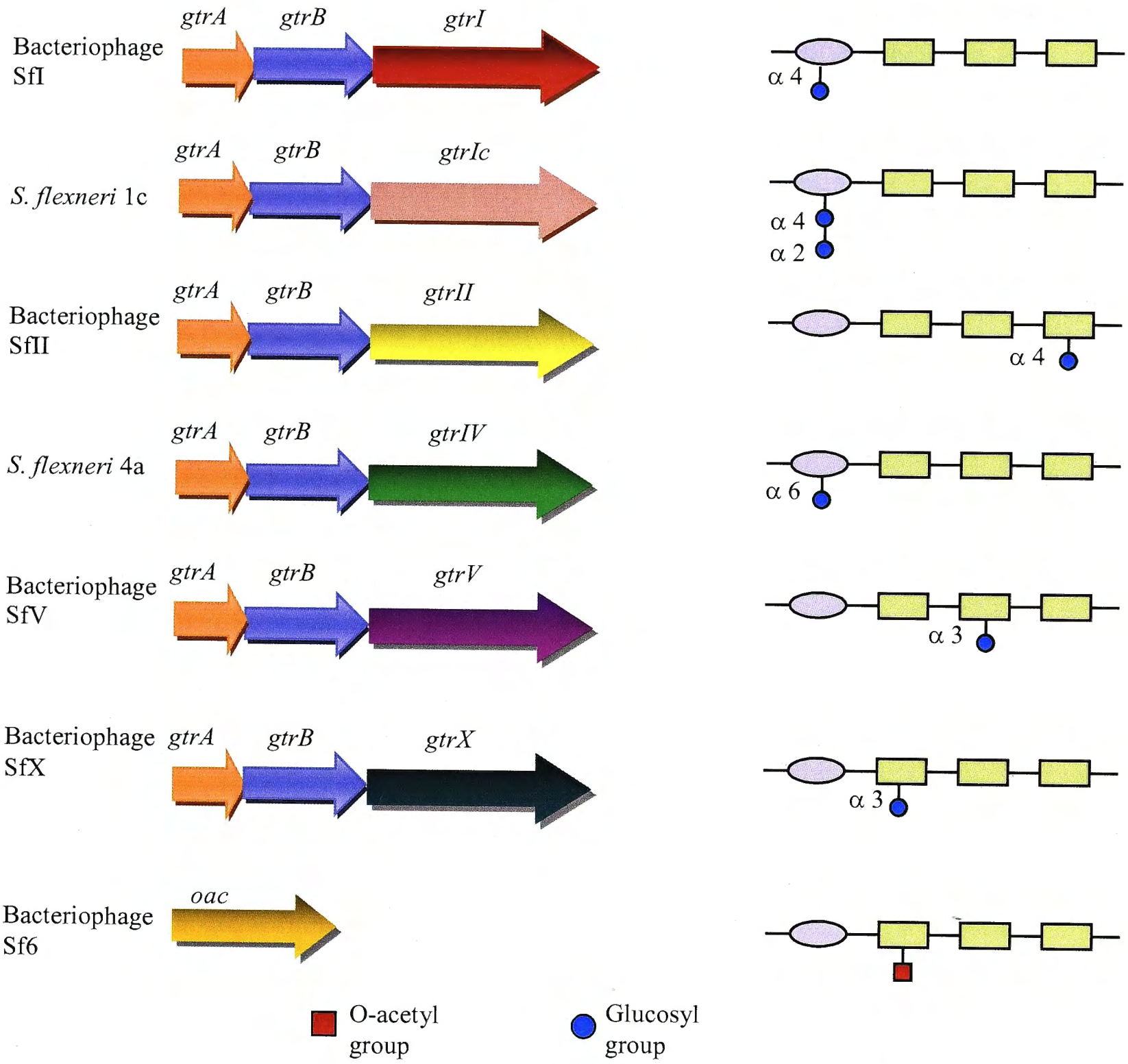


Figure 1.10: Organisation of *S. flexneri* O-antigen modification genes. The glucosylation cassette comprises of three genes: *gtrA*, *gtrB* and *gtr(type)*. *gtrA* and *gtrB* are conserved among serotypes while *gtr(type)* is unique and serotype specific. Bacteriophage Sf6 uses only a single gene. This gene, *oac*, is responsible for acetylation of O-antigens. No bacteriophage for *S. flexneri* 1c and *S. flexneri* 4a has yet been isolated. The O-antigen modification mediated by each cluster of genes is shown on the right. Adapted from H. Korres (PhD Thesis, 2006) and V. Tran (Honours Thesis, 2010)

specificity. However, no direct evidence has yet been put forward to support this. The GtrB proteins, encoded by *gtrB,* vary in size from 305 to 309 amino acids. As with GtrA, GtrB is highly conserved between the *S. flexneri* serotypes. Hydropathy data suggests that GtrB has two transmembrane domains with a large hydrophilic N-terminus and a short C-terminal tail (Korres *et al.*, 2005). GtrB catalyses the transfer of glucose from UDP-Glucose to bactoprenol phosphate to form UndP-β-glucose in the cytoplasm (Guan *et al.*, 1999). This structure is then flipped by GtrA into the periplasm before the glucosyl residue is attached by the $Gtr_{(type)}$ to the growing O-antigen unit (Korres *et al.*, 2005). GtrB homologues are found in a number of organisms, including *E. coli* and *Salmonella* (Mavris *et al.*, 1997).

The $gtr_{(type)}$ is the third gene in the three-gene glucosylation cluster involved in serotype conversion. It encodes for a serotype-specific glucosyltransferase (Gtr) which is thought to be responsible for the actual attachment of glucosyl residues onto the appropriate sugar residue of the O-antigen repeat unit (Guan *et al.*, 1999). Till now, the various $Gtr_{(type)}$ proteins that have been identified vary in size from 416 to 526 residues and have 9-11 putative transmembrane regions. The secondary structure of the different $Gtr_{(type)}$ proteins is similar to flipases, O-antigen polymerases and O-antigen ligases (Whitfield, 1995) As such, it has been proposed that, apart from attaching a glucosyl residue to the appropriate sugar residue of the O-antigen, Gtr(type) may also play a role in recycling the lipid carrier after addition of the glucosyl group to the O-antigen (Guan *et al.*, 1999). Figure 1.11 summarises the proposed mechanism for O-antigen glucosylation.

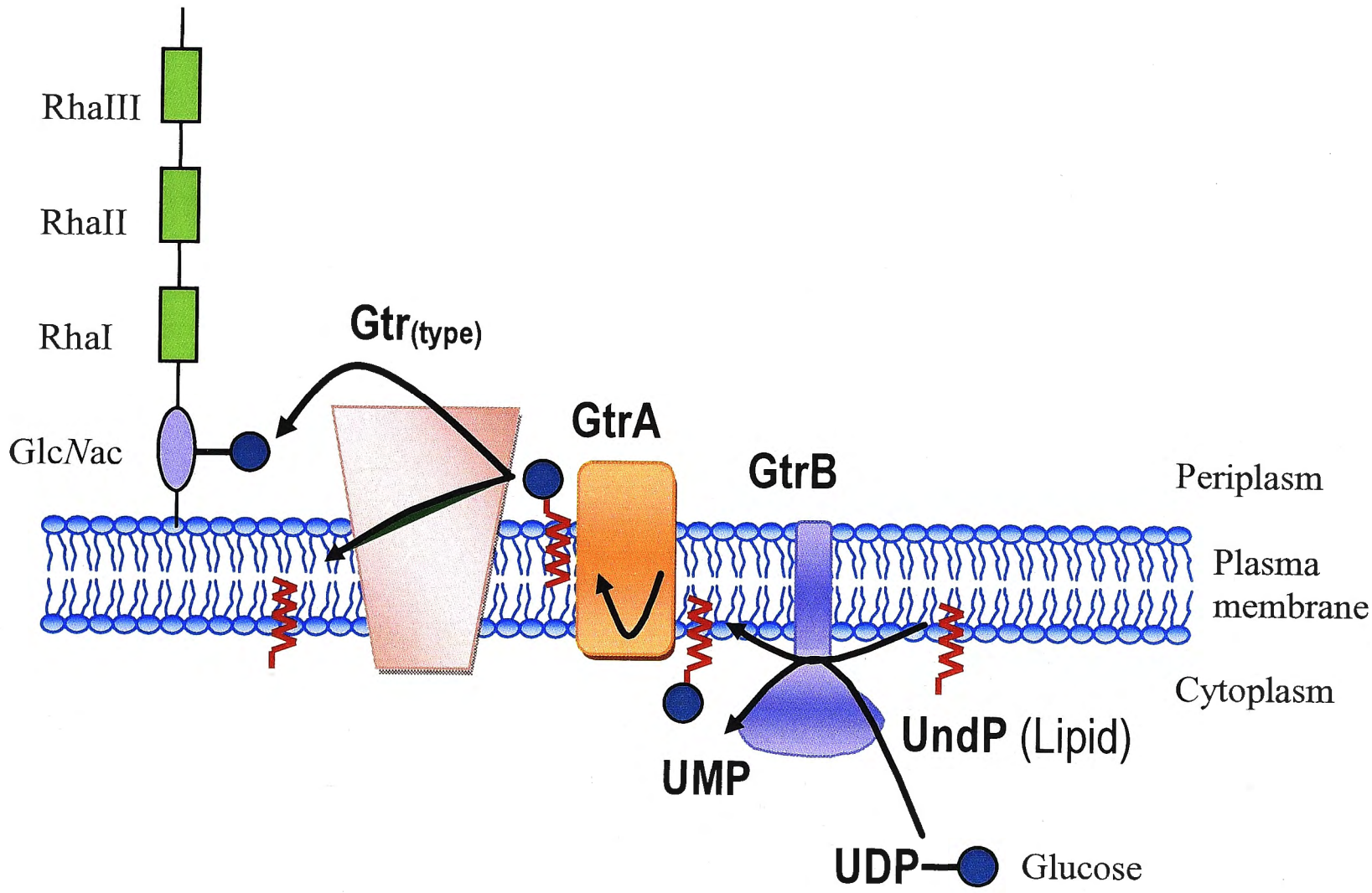


Figure 1.11: The hypothetical model of the glucosylation of the O-antigen. GtrB transfers a glucosyl group from UDP-glucose to the membrane linked-lipid UndP. The GtrA protein itself or in conjunction with the $Gtr_{(type)}$, flips the lipid associated glucose into the periplasm. The $Gtr_{(type)}$ protein then specifically transfers the glucosyl residue to a particular site on the O-antigen chain and returns the bactoprenol to the cytoplasmic side of the membrane. Glc*N*Ac, N-acetylglucosamine; Rha, Rhamnose; UDP, uridine diphosphate; UndP, undecaprenyl phosphate. Adapted from Guan *et al.,* (1999) and Brockhausen *et al.,* (2008)

1.5.4.3 O-acetylation of the O-antigen

O-acetylation of the O-antigen is carried out by the enzyme, O-acetyltransferase (Oac). The gene encoding for the enzyme differs from the *gtr* cassette in that it lies in an independent open reading frame (Ramiscal *et al.,* 2010). Oac is able to convert *S. flexneri* serotype Y, 1a, 4a to 3b, 1b, 4b, respectively, by adding an acetyl group to the sugar residues of the O-antigen (Allison & Verma, 2000). Unfortunately, it is still not known where O-acetylation of the O-antigen takes place within the cell.

1.5.5 Emergence of new serotypes

In recent times, it was reported new *S. flexneri* serotypes such as serotype 1c and SFxv had emerged. Serotype 1c was first identified in Bangladesh in 1989 (Carlin N. I. *et al,.* 1989) and was later found to be prevalent in Bangladesh, Egypt, Indonesia, and Pakistan (El-Gendy *et al.,* 1999; Talukder *et al.,* 2002; von Seidlein *et al.,* 2006). At present, it is the most prevalent serotype in Vietnam (Stagg *et al.,* 2008). SFxv first appeared in Henan province in 2001 and spread to several other provinces, including Shanxi, Gansu, Anhui, and Shanghai and has become one of the predominant serotypes in these Chinese provinces (Ye *et al.,* 2010; Sun *et al,* 2011; Qiu *et al.,* 2011). Less than a year ago, Qiu *et al.,* (2011) described the first isolation of a new *S. flexneri* serotype, designated 4s, in Beijing, China. With the emergence of untypeable or novel serotypes of *S. flexneri* from natural infections, Sun *et al.,* (2011) sought to investigate the roles played by serotype-converting phages in the emergence of new serotypes and also the potential of emergence of novel serotypes through this mechanism in nature. In doing so, they successfully engineered a new *S. flexneri* serotype that was termed serotype 1d by infecting an *S. flexneri* serotype Y strain (native LPS) sequentially with 2 serotype-converting bacteriophages,

SfX first and then SfI. This lab-generated strain displayed slide agglutination characteristics of both serotype X and serotype 1a.

1.5.6 Distribution of Shigella flexneri *serotypes*

Through the use of serotyping techniques, aetiological studies found that *S. flexneri* serotype 2a was the most predominant serotype in both developing and developed countries (Kotloff *et al.,* 1999; Ram *et al.,* 2008). It was further found that serotype 1a, 1b, 1c, 2a, 3a, 4a and 6 are commonly isolated in South and East Asian countries (Kotloff *et al.,* 1999; Ram *et al.,* 2008; Stagg *et al.,* 2008). This pattern changes with area as well as time, with new serotypes and untypeable strains rising frequently. Ahmed *et al,* (2006) reported that one of the newly identified serotypes, *S. flexneri* 1c, has emerged as a dominant *S. flexneri* serotype in Egypt. A study by Stagg *et al.,* (2008) showed *S. flexneri* serotype 1c to be the most prevalent serotype within the Son Tay Province, Vietnam. The authors went on to suggest that there is a large variation of *S. flexneri* serotypes within different regions of Vietnam (Stagg *et al.*, 2008). Alarmingly, reports of new serotypes 4s and SFxv have surfaced within several provinces in China. (Ye *et al.,* 2010; Sun *et al,* 2011; Qiu *et al.,* 2011)

1.6 Structural and functional studies of O-antigen modifying enzymes

A major drawback with current live-attenuated vaccines is that they are serotype specific. Therefore, to create a multi-valent vaccine for shigellosis, the mechanisms behind how Gtrs and Oac modify the O-antigen need to be determined. By identifying the structures of the enzymes, a better understanding of how the enzymes interact with one another is provided. Work on

glycosyltransferases (GTs) in the past has helped shed light on these integral membrane proteins that recognise identical donor or acceptor substrates. Firstly, it was discovered that few regions of sequence homology exist. It was also shown that enzymes that are structurally related often catalyse the same or a similar reaction (Breton and Imberty, 1999). Although sequence-based classification places GTs in many families, the 3-D structures of glycosyltransferases have revealed two major structural folds that can be grouped into 2 super families GT-A and GT-B (Breton and Imberty, 1999; Lairson *et al.,* 2008). *S. flexneri* glucosyltransferases bear similarity to, but are not part of the GT-A superfamily although they contain a DxD motif similar to that witnessed in the GT-A superfamily (Breton and Imberty, 1999; Lairson *et al.,* 2008). Therefore, by gathering as much structural and functional information as we can about Gtrs, it will enable us to make comparisons with the large library of GTs in order to elucidate their mode of action and determine their catalytic domains.

Lehane *et al.* (2005) revealed that GtrII has a cytoplasmic N-terminus, nine transmembrane helices, a re-entrant loop and a large periplasmic C-terminal region (Lehane *et al.*, 2005). The enzyme's hydropathy profile was found to be similar to GtrI, indicating that there is some structural homology between the two $Gtr_{(type)}$ proteins (Lehane *et al.*, 2005). Four critical residues were identified in GtrII (D40, F414, C435 and L478) (Lehane *et al.*, 2005). C435 forms a disulfide bond with C437 that was postulated to be required for the positioning of other critical residues (Lehane *et al.*, 2005). D40 was hypothesised to be critical as, many Gtrs use either glutamic or aspartic acid residues to help deprotonate the hydroxyl group of the acceptor sugar (Lehane *et al.*, 2005). In the same vein, GtrV was experimentally found to have a structure consisting of a cytoplasmic N-terminus, nine transmembrane helices, a re-entrant loop and a

large periplasmic C-terminal region. The same study also found that GtrV had strong sequence similarity to GtrX, and that there were three motifs (two in loop No. 2 and one in loop No. 1), which contain acidic residues critical to GtrV function (Korres & Verma, 2006). GtrIc is the largest member of the Gtr family and is 526 amino acids in size. GtrIc consists of ten transmembrane helices, cytoplasmic N- and C-termini, two large periplasmic loops and a double intramembrane dipping loop (Ramiscal *et al.*, 2010). Even though the topology of GtrIc is unique comparative to other members of the Gtr family, this protein shares some structural homology with the predicted GtrIV model as both proteins have cytoplasmic N- and C-termini along with two large periplasmic loops. Figure 1.12 illustrates compares the known topologies of all the Gtrs to date.

Studies on Oac showed that the enzyme is an integral membrane protein consisting of ten transmembrane helices, a cytoplasmic N-terminus and a periplasmic C-terminus (Thanweer *et al.*, 2008). Two domains (one in loop No. 3 and the other in loop No. 4) were found to be critical, as deletion studies resulted in a loss of function (Thanweer *et al.*, 2008). Through site-directed mutagenesis, the study discovered three critical residues (R73 and R75-R76) in loop No. 3 that are critical to the enzymes function. However, the role of these residues has yet to be determined (Thanweer *et al.*, 2008).

The construction of chimeric proteins is a way in which investigations can be carried out into Chimeric experiments have been conducted on members of the Gtr family to determine if any regions carry out conserved functions between the Gtr proteins. The second purpose was to discover whether a chimera protein with the capability to convert two or more serotypes could be

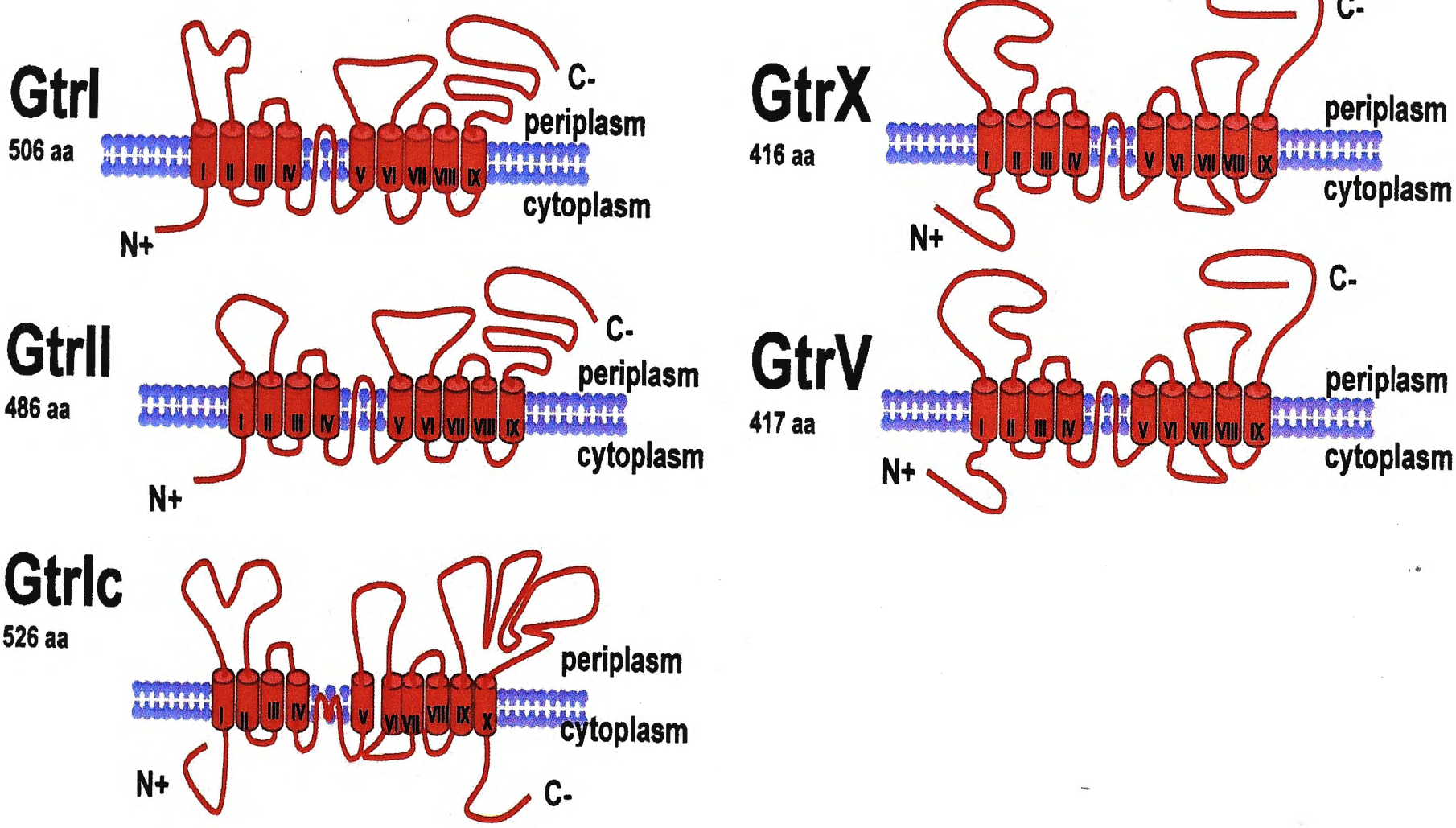


Figure 1.12: Topological comparison of all the Gtrs from the *S. flexneri* family. A conserved reentrant loop is present in GtrI, GtrII, GtrV and GtrX. Also Loop No. 2 is conserved indicating a conserved mode of action. Extended periplsmic loops are also evident in GtrI and GtrII in the C-terminal end. Adapted from V. Tran (Honours thesis, 2010)

created. Previously in our lab, given the structural homology between GtrV and GtrX, chimeric protein experiments were conducted between the two proteins (Korres & Verma, 2006). It was found that when loop No. 2 of GtrX was swapped into GtrV, the resulting chimera still retained its function. This led the authors to hypothesize that the N-terminal periplasmic regions of the Gtrs had a conserved function.

1.7 Main objectives

In order to develop a vaccine strain expressing multiple serotypes a number of process have to be understood. One of these processes is the glucosylation of the O-antigen conferring serotype specificity. As seen in the introduction, the mechanism involves three glucosyltransferases GtrA, GtrB and $Gtr_{(type)}$. The broad aim of this project was to gain insight into the structures and mechanisms of action of the *S. flexneri* glucosyltransferase IV (GtrIV), which is encoded by the cryptic bacteriophage SfIV. This may contribute to the ultimate goal of developing a polyvalent vaccine strain expressing multiple Gtrs. GtrIV adds a glucosyl residue to *N*-acetylglucosamine of the O-antigen repeat unit via an α1,6 linkage, thus converting serotype Y to the virulent serotype 4a. The main aims of this project are as follows:

1. Determine the topology of GtrIV

The topological information gathered from this objective will be useful for the identification of putative catalytic domains and residues critical for GtrIV function that may reside within these domains.

2. ***Investigate the functions of the large periplasmic loops of GtrIV***

By determining whether the periplasmic loops of GtrIV are required for catalytic activity or to maintain structural stability of the protein, it will help narrow down segments within the loops that may contain potential catalytic sites. This will be achieved through sequential loop deletion experiments. Through the creation of chimeric proteins between GtrIV and its closest structural homologue GtrIc, regions that confer their serotype specificity can be identified. The localisation of specific functional domains within GtrIV would accelerate the identification of critical residues.

3. ***Locate critical residues in GtrIV***

In conjunction with topological information and the localisation of distinct domains, the identification of critical residues would enable the development of a model mechanism of action of GtrIV, which may be unique as compared to the other Gtrs.

4. ***Gain insight into genetic diversity of wildtype serotype 4a strains***

By carrying out sequence analysis and Southern blots on the various wildtype serotype 4a strains isolated from Japan, Bangladesh and Vietnam, we will be able to determine if there is any genetic variation between the strains of each region. This can help give an insight into the evolution of the serotype 4a strains and may help uncover the presence of the cryptic bacteriophage SfIV.

Chapter 2

Materials & Methods

All centrifugation steps were performed at 16,000 x g unless otherwise stated. For reagents that needed to be made sterile, they were either autoclaved at 121°C for 15 minutes or filter sterilized through 0.2 μm filter (Minisart).

2.1 Bacterial Strains and Growth Conditions

All bacterial cultures were grown aerobically at 37°C on either Luria-Bertani (LB) agar plates (Appendix A) or in LB medium (Appendix A). Cultures were incubated with shaking (150-200 rpm). Antibiotics such as chloramphenicol (Cm), kanamycin (Kan) and ampicillin (Amp) were added to media when required. The final concentrations of these antibiotics were 25 μg/mL, 50 μg/mL, and 100 μg/mL, respectively. Bacterial plates were stored for up to 3 weeks at 4°C in the cold room. For long term storage, bacterial strains were preserved in 1 mL of glycerol:LB (1:1 v/v). The glycerol stocks were then stored at -80°C.

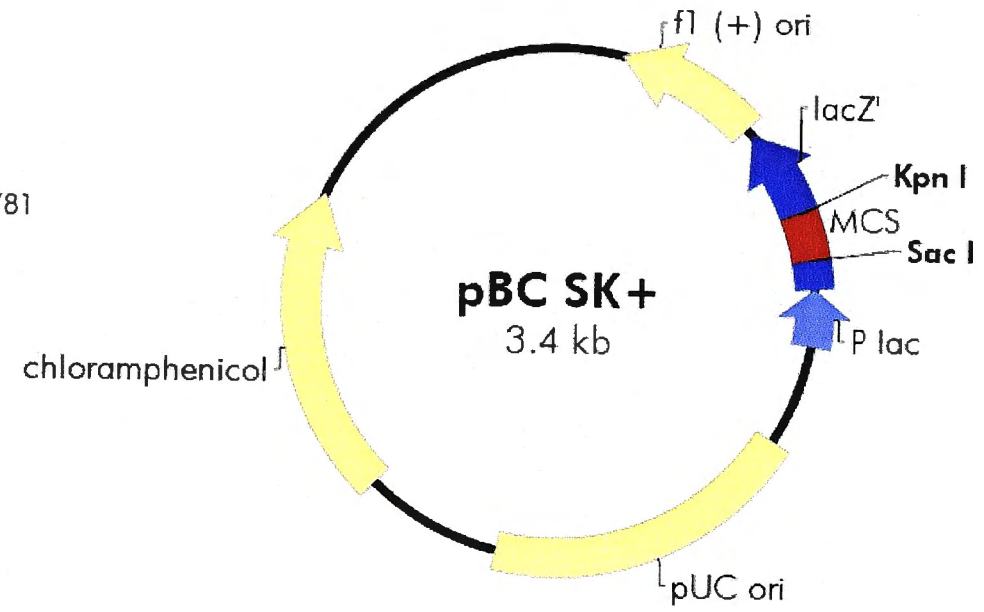
The *E. coli* strains used in this study are derivatives of *E. coli* K-12. JM109 cells were used for topology studies while XL1-Blue cells for routine cloning procedures. The *E. coli* and *S. flexneri* strains used in this study are listed in Tables 2.1 and 2.2, respectively.

2.2 DNA Preparation

2.2.1 Plasmids

The plasmids used and constructed in this study are shown in Table 2.3. All the plasmids constructed were derived from the cloning vector pBC SK (Stratagene) and is shown in Figure 2.1.

f1 (+) origin 135–441
β-galactosidase α-fragment 460–816
multiple cloning site 653–760
***lac* promoter** 817–938
pUC origin 1158–1825
chloramphenicol resistance ORF 2125–2781



pBC SK (+/–) Multiple Cloning Site Region
(sequence shown 598–826)

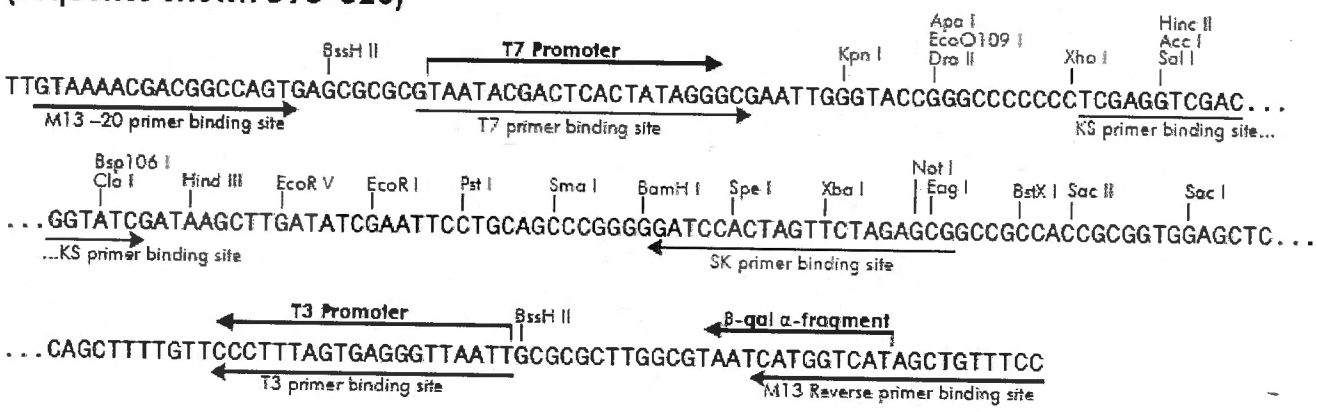


Figure 2.1: Map of cloning vector pBC SK+ (adapted from Stratagene)

Table 2.1: *E. coli* strains used and created in this study.

Strain	Description of Genotype	Source
JM109	*rec*A1 *sup*E44 *end*A1 *hsd*R17 *gyr*A96 *rel*A1 *thi* Δ(*lac-pro*AB) [F' *tra*D36 *pro*AB *lac*I^q *lac*ZΔM15]	Yanisch-Perron *et al.* (1985)
XL1-Blue MRF'	*sup*E44 *hsd*R17 *rec*A1 *end*A1 *gyr*A96 *thi*-1 *rel*A1 *lac* [F' *pro*AB *lac*I^{q}ZΔM15 *T*n10 (Tet^r)]	Jerpseth *et al.* (1992)
B1758	JM109 carrying pNV1473	Nair *et al.*, (2011)
B1904	JM109 carrying pNV1562	Nair *et al.*, (2011)
B1905	JM109 carrying pNV1563	Nair *et al.*, (2011)
B1906	JM109 carrying pNV1564	Nair *et al.*, (2011)
B1907	JM109 carrying pNV1565	Nair *et al.*, (2011)
B1908	JM109 carrying pNV1566	Nair *et al.*, (2011)
B1909	JM109 carrying pNV1567	Nair *et al.*, (2011)
B1910	JM109 carrying pNV1568	Nair *et al.*, (2011)
B1911	JM109 carrying pNV1569	Nair *et al.*, (2011)
B1912	JM109 carrying pNV1570	Nair *et al.*, (2011)
B1913	JM109 carrying pNV1571	Nair *et al.*, (2011)
B1914	JM109 carrying pNV1572	Nair *et al.*, (2011)
B1919	JM109 carrying pNV1577	Nair *et al.*, (2011)
B1920	JM109 carrying pNV1578	Nair *et al.*, (2011)
B1921	JM109 carrying pNV1579	Nair *et al.*, (2011)
B1922	JM109 carrying pNV1580	Nair *et al.*, (2011)
B1923	JM109 carrying pNV1581	Nair *et al.*, (2011)
B1924	JM109 carrying pNV1582	Nair *et al.*, (2011)
B1925	JM109 carrying pNV1583	Nair *et al.*, (2011)
B1926	JM109 carrying pNV1584	Nair *et al.*, (2011)
B1933	JM109 carrying pNV1591	Nair *et al.*, (2011)
B1934	JM109 carrying pNV1592	Nair *et al.*, (2011)
B1935	JM109 carrying pNV1593	Nair *et al.*, (2011)
B1936	JM109 carrying pNV1594	Nair *et al.*, (2011)
B1937	JM109 carrying pNV1595	Nair *et al.*, (2011)
B2182	JM109 carrying pNV1792	Ramiscal *et al.*, (2010)
B2338	JM109 carrying pNV1920	This study
B2339	JM109 carrying pNV1917	This study
B2340	JM109 carrying pNV1921	This study
B2341	JM109 carrying pNV1924	This study
B2342	JM109 carrying pNV1926	This study
B2343	JM109 carrying pNV1922	This study
B2344	JM109 carrying pNV1927	This study
B2345	JM109 carrying pNV1925	This study
B2346	JM109 carrying pNV1923	This study
B2347	XL1-Blue MRF' carrying pNV1918	This study
B2348	XL1-Blue MRF' carrying pNV1919	This study
B2349	XL1-Blue MRF' carrying pNV1929	This study
B2350	XL1-Blue MRF' carrying pNV1928	This study
B2351	XL1-Blue MRF' carrying pNV1931	This study
B2352	XL1-Blue MRF' carrying pNV1930	This study
B2353	XL1-Blue MRF' carrying pNV1932	This study

Table 2.1 continued

B2354	XL1-Blue MRF' carrying pNV1933	This study
B2355	XL1-Blue MRF' carrying pNV1934	This study
B2356	XL1-Blue MRF' carrying pNV1935	This study
B2357	XL1-Blue MRF' carrying pNV1936	This study
B2358	XL1-Blue MRF' carrying pNV1937	This study
B2359	XL1-Blue MRF' carrying pNV1938	This study
B2360	XL1-Blue MRF' carrying pNV1939	This study
B2443	XL1-Blue MRF' carrying pNV1998	This study
B2444	XL1-Blue MRF' carrying pNV1999	This study
B2445	XL1-Blue MRF' carrying pNV2000	This study
B2446	XL1-Blue MRF' carrying pNV2001	This study
B2447	XL1-Blue MRF' carrying pNV2002	This study
B2448	XL1-Blue MRF' carrying pNV2003	This study
B2449	XL1-Blue MRF' carrying pNV2004	This study
B2450	XL1-Blue MRF' carrying pNV2005	This study
B2451	XL1-Blue MRF' carrying pNV2006	This study
B2452	XL1-Blue MRF' carrying pNV2007	This study
B2453	XL1-Blue MRF' carrying pNV2008	This study
B2454	XL1-Blue MRF' carrying pNV2009	This study
B2455	XL1-Blue MRF' carrying pNV2010	This study
B2456	XL1-Blue MRF' carrying pNV2011	This study
B2457	XL1-Blue MRF' carrying pNV2012	This study
B2458	XL1-Blue MRF' carrying pNV2013	This study
B2459	XL1-Blue MRF' carrying pNV2014	This study
B2460	XL1-Blue MRF' carrying pNV2015	This study
B2461	XL1-Blue MRF' carrying pNV2016	This study
B2462	XL1-Blue MRF' carrying pNV2017	This study
B2463	XL1-Blue MRF' carrying pNV2018	This study
B2464	XL1-Blue MRF' carrying pNV2019	This study
B2465	XL1-Blue MRF' carrying pNV2020	This study

Table 2.2: *S. flexneri* strains used and created in this study.

Strain	Description of Genotype	Source
SFL1253	Wildtype *S. flexneri* serotype 4a	NCTC 8296
SFL1255	Wildtype *S. flexneri* serotype 4b	NCTC 8336
SFL1294	Wildtype *S. flexneri* serotype 4a	C. Sasakawa, Japan
SFL1295	Wildtype *S. flexneri* serotype 4a	C. Sasakawa, Japan
SFL1304	Wildtype *S. flexneri* serotype 4a	C. Sasakawa, Japan
SFL1305	Wildtype *S. flexneri* serotype 4b	C. Sasakawa, Japan
SFL1314	Wildtype *S. flexneri* serotype 4a	C. Sasakawa, Japan
SFL1328	Wildtype *S. flexneri* serotype 4a	C. Sasakawa, Japan
SFL1416	Wildtype S. flexneri serotype 1a	NCTC
SFL1522	Wildtype *S. flexneri* serotype 4a	ICDDR, Bangladesh
SFL1523	Wildtype *S. flexneri* serotype 4a	ICDDR, Bangladesh
SFL1524	Wildtype *S. flexneri* serotype 4a	ICDDR, Bangladesh
SFL1525	Wildtype *S. flexneri* serotype 4a	ICDDR, Bangladesh
SFL1526	Wildtype *S. flexneri* serotype 4a	ICDDR, Bangladesh
SFL1616	SFL124 with *gtrA* and *gtrB* from SfV in pACYC177.	Lehane *et al.,* (2005)
SFL1714	Wildtype *S. flexneri* serotype 4a	P. D. Cam, Vietnam
SFL1758	Wildtype *S. flexneri* serotype 4	P. D. Cam, Vietnam
SFL1767	Wildtype *S. flexneri* serotype 4	P. D. Cam, Vietnam
SFL1768	Wildtype *S. flexneri* serotype 4	P. D. Cam, Vietnam
SFL1769	Wildtype *S. flexneri* serotype 4	P. D. Cam, Vietnam
SFL2069	Wildtype serotype 1a host to pNV1792	Ramiscal *et al.,* (2011)
SFL2175	Wildtype *S. flexneri* serotype 4a	P. D. Cam, Vietnam
SFL2176	Wildtype *S. flexneri* serotype 4a	P. D. Cam, Vietnam
SFL2178	Wildtype *S. flexneri* serotype 4a	P. D. Cam, Vietnam
SFL2179	Wildtype *S. flexneri* serotype 4a	P. D. Cam, Vietnam
SFL2180	Wildtype *S. flexneri* serotype 4a	P. D. Cam, Vietnam
SFL2183	Wildtype *S. flexneri* serotype 4	P. D. Cam, Vietnam
SFL2187	Wildtype *S. flexneri* serotype 4a	P. D. Cam, Vietnam
SFL2193	Wildtype *S. flexneri* serotype 4a	P. D. Cam, Vietnam
SFL2234	Wildtype *S. flexneri* serotype 4	P. D. Cam, Vietnam
SFL2241	Wildtype *S. flexneri* serotype 4	P. D. Cam, Vietnam
SFL2265	Wildtype *S. flexneri* serotype 4a	P. D. Cam, Vietnam
SFL2271	Wildtype *S. flexneri* serotype 4a	P. D. Cam, Vietnam
SFL2324	SFL1616 carrying pNV1917	This study
SFL2325	SFL1616 carrying pNV1918	This study
SFL2326	SFL1616 carrying pNV1919	This study
SFL2327	SFL1616 carrying pNV1929	This study
SFL2328	SFL1616 carrying pNV1928	This study
SFL2329	SFL1616 carrying pNV1931	This study
SFL2330	SFL1616 carrying pNV1930	This study
SFL2331	SFL1616 carrying pNV1932	This study
SFL2332	SFL1616 carrying pNV1933	This study
SFL2333	SFL1616 carrying pNV1934	This study
SFL2334	SFL1616 carrying pNV1935	This study
SFL2335	SFL1616 carrying pNV1936	This study
SFL2336	SFL1616 carrying pNV1937	This study

Table 2.2 continued

SFL2337	SFL1616 carrying pNV1938	This study
SFL2338	SFL1616 carrying pNV1939	This study
SFL2339	SFL1416 carrying pNV1936	This study
SFL2340	SFL1416 carrying pNV1937	This study
SFL2341	SFL1416 carrying pNV1938	This study
SFL2342	SFL1416 carrying pNV1939	This study
SFL2388	SFL1616 carrying pNV1998	This study
SFL2389	SFL1616 carrying pNV1999	This study
SFL2390	SFL1616 carrying pNV2000	This study
SFL2391	SFL1616 carrying pNV2001	This study
SFL2392	SFL1616 carrying pNV2002	This study
SFL2393	SFL1616 carrying pNV2003	This study
SFL2394	SFL1616 carrying pNV2004	This study
SFL2395	SFL1616 carrying pNV2005	This study
SFL2396	SFL1616 carrying pNV2006	This study
SFL2397	SFL1616 carrying pNV2007	This study
SFL2398	SFL1616 carrying pNV2008	This study
SFL2399	SFL1616 carrying pNV2009	This study
SFL2400	SFL1616 carrying pNV2010	This study
SFL2401	SFL1616 carrying pNV2011	This study
SFL2402	SFL1616 carrying pNV2012	This study
SFL2403	SFL1616 carrying pNV2013	This study
SFL2404	SFL1616 carrying pNV2014	This study
SFL2405	SFL1616 carrying pNV2015	This study
SFL2406	SFL1616 carrying pNV2016	This study
SFL2407	SFL1616 carrying pNV2017	This study
SFL2408	SFL1616 carrying pNV2018	This study
SFL2409	SFL1616 carrying pNV2019	This study
SFL2410	SFL1616 carrying pNV2020	This study

ICDDR – International Centre for Diarrhoeal Disease Research

Table 2.3: Plasmids used and constructed in this study.

Plasmid	Relevant Characteristics	Source
pBC SK	3.4 kb cloning vector (Fig. 2.1)	Stratagene
pNV1473	Product of ligation between *gtrIV* and majority of pNV1090 (excluding *gtrV*) (*Nhe*I/*Xba*I)	Nair *et al.*, (2011)
pNV1792	pNV1734 having *gtrIc* downstream terminus fused to *phoA/lacZ*	Ramiscal *et al.*, (2010)
pNV1914	pNV1473 with *phoA/lacZ* excised out	This study
pNV1917	pNV1473 with *phoA/LacZ* fused to C-terminus of *gtrIV*	This study
pNV1918	pNV1917 with GtrIV loop No. 2 deletion. Contains *Sma*I site at point of deletion	This study
pNV1919	pNV1917 with GtrIV loop No. 3 deletion. Contains *Sma*I site at point of deletion	This study
pNV1920	pNV1473 in-frame PCR-based fusion with PhoA/LacZ at residue N255	This study
pNV1921	pNV1473 in-frame PCR-based fusion with PhoA/LacZ at residue V102	This study
pNV1922	pNV1473 in-frame PCR-based fusion with PhoA/LacZ at residue V155	This study
pNV1923	pNV1473 in-frame PCR-based fusion with PhoA/LacZ at residue D169	This study
pNV1924	pNV1914 with PhoA/LacZ sandwich fusion at residue R93	This study
pNV1925	pNV1914 with PhoA/LacZ sandwich fusion at residue K117	This study
pNV1926	pNV1914 with PhoA/LacZ sandwich fusion at residue A181	This study
pNV1927	pNV1914 with PhoA/LacZ sandwich fusion at residue R375	This study
pNV1928	pNV1917 with GtrIV loop No. 6 deletion. Contains *Sma*I site at point of deletion	This study
pNV1929	pNV1917 with GtrIV loop No. 6 partial deletion 1. Contains *Sma*I site at point of deletion	This study
pNV1930	pNV1917 with GtrIV loop No. 6 partial deletion 2. Contains *Sma*I site at point of deletion	This study
pNV1931	pNV1917 with GtrIV loop No. 6 partial deletion 3. Contains *Sma*I site at point of deletion	This study
pNV1932	pNV1917 with GtrIV deletion of residues N251-N255. Contains *Sma*I site at point of deletion	This study
pNV1933	pNV1917 with GtrIV deletion of residues P256-V260. Contains *Sma*I site at point of deletion	This study
pNV1934	pNV1917 with GtrIV deletion of residues D261-G265. Contains *Sma*I site at point of deletion	This study
pNV1935	pNV1917 with GtrIV deletion of residues V266-W269. Contains *Sma*I site at point of deletion	This study
pNV1936	GtrIc-GtrIV Loop No. 2-GtrIc chimera. *Bgl*II and *Mlu*I sites at points of fusion	This study
pNV1937	GtrIV-GtrIc Loop No. 2-GtrIV chimera. *Bgl*II and *Mlu*I sites at points of fusion	This study
pNV1938	GtrIc-GtrIV Loop No. 6-GtrIc chimera. *Bgl*II and *Mlu*I sites at points of fusion	This study
pNV1939	GtrIV-GtrIc Loop No. 10-GtrIV chimera. *Bgl*II and *Mlu*I sites at points of fusion	This study
pNV1973	pNV1914 with PhoA/LacZ sandwich fusion at residue K162	This study

Table 2.3 continued

pNV2000	pNV1917 after site-directed mutagenesis to E67 mutated to A in GtrIV	This study
pNV2001	pNV1917 after site-directed mutagenesis to DE261 mutated to AA in GtrIV	This study
pNV2002	pNV1917 after site-directed mutagenesis to ED326 mutated to AA in GtrIV	This study
pNV2003	pNV1917 after site-directed mutagenesis to ED283 mutated to AA in GtrIV	This study
pNV2004	pNV1917 after site-directed mutagenesis to D267 mutated to A in GtrIV	This study
pNV2005	pNV1917 after site-directed mutagenesis to E254 mutated to A in GtrIV	This study
pNV2006	pNV1917 after site-directed mutagenesis to E58 mutated to A in GtrIV	This study
pNV2007	pNV1917 after site-directed mutagenesis to E333 mutated to A in GtrIV	This study
pNV2008	pNV1917 after site-directed mutagenesis to E294 mutated to A in GtrIV	This study
pNV2009	pNV1917 after site-directed mutagenesis to D274 mutated to A in GtrIV	This study
pNV2010	pNV1917 after site-directed mutagenesis to D49 mutated to A in GtrIV	This study
pNV2011	pNV2005 after site-directed mutagenesis to D267 mutated to A in GtrIV	This study
pNV2012	pNV2001 after site-directed mutagenesis to E254 mutated to A in GtrIV	This study
pNV2013	pNV2001 after site-directed mutagenesis to D267 mutated to A in GtrIV	This study
pNV2014	pNV1917 after site-directed mutagenesis to W269 mutated to A in GtrIV	This study
pNV2015	pNV1917 after site-directed mutagenesis to I264 mutated to A in GtrIV	This study
pNV2016	pNV1917 after site-directed mutagenesis to V266 mutated to A in GtrIV	This study
pNV2017	pNV1917 after site-directed mutagenesis to G265 mutated to A in GtrIV	This study
pNV2018	pNV2013 after site-directed mutagenesis to A262 mutated to E in GtrIV	This study
pNV2019	pNV2013 after site-directed mutagenesis to A261 mutated to D in GtrIV	This study
pNV2020	pNV1917 after site-directed mutagenesis to C263 mutated to A in GtrIV	This study

2.2.2 Isolation of plasmid DNA

When plasmid DNA was required for either sequencing or cloning, extraction of plasmids was carried out using Axygen Axyprep Plasmid Miniprep Kit. A 1.5 mL aliquot of a 5 mL overnight culture was transferred to a 1.5 mL Eppendorf tube and centrifuged for 2 min, before removing the supernatant. The resulting pellet was resupended in 250 μL of Buffer S1. 250 μL of Buffer S2 was added to the tube and inverted 4-6 times. 350 μL of Buffer S3 was then added and the tube was inverted for 6-8 times before being centrifuged for 10 min. The supernatant was transferred to a Miniprep column assembly and centrifuged for 1 min. The flow-through solution was discarded, and the column was washed with 500 μL of Buffer W1, followed by two washes with 700 μL of Buffer W2. After discarding the flow-through solution, the MiniPrep column assembly was then centrifuged for 1 minute to remove any excess flow-through solution. The AxyPrep column was then placed in a sterile 1.5 mL Eppendorf tube and 60 μL of eluent was added to the centre of the column. The column was allowed to stand for 1 minute before being centrifuged for 1 minute to elute the plasmid DNA.

When plasmid DNA needed to be isolated from many cultures for screening purposes, the alkaline lysis method adapted by Sambrook & Russell (2001) from Birnboim & Doly (1979). Bacteria were pelleted as described above. Once the supernatant medium was discarded, the pellet was resuspended in 100 μL of ice-cold Alkaline Lysis Solution I (Appendix A). 200 μL of freshly made Alkaline Lysis Solution II (Appendix A) was then added and the tube was inverted 5-10 times and placed on ice. 150 μL of ice-cold Alkaline Lysis Solution III (Appendix A) was then added and the tube was inverted immediately but gently 5-10 times and left on ice for 5 min. The tube was then centrifuged for 5 minutes and the supernatant solution was transferred into a clean microcentrifuge tube.

The plasmid DNA was then precipitated by adding 2 volumes of absolute ethanol. The tube was inverted and allowed to stand for 2 min. Following a 5 minutes centrifugation, the supernatant was discarded and 300 µL of 70% ethanol was added. The tube was vortexed and centrifuged for 2 min. The supernatant was discarded and the plasmid DNA was dried under vacuum (Savant SpeedVac SC100 and high drying rate) for 10 min. The plasmid DNA was resuspended in sterile MilliQ water (typically 50 µL).

2.2.3 Genomic DNA extraction

Genomic DNA was purified from the *S. flexneri* wildtype 4 strains using the Illustra™ Bacteria GenomicPrep Mini Spin Kit (GE). 1 mL of overnight culture was centrifuged for 30 seconds at 16,000 x g. After discarding the supernatant, 40 µL of Lysis buffer type 2, 10 µL of proteinase K and 10 µL of Lysis buffer type 3 was added and the tube incubated at 55°C for 7 minutes. After which, the tube was vortexed, centrifuged at 1,000 x g for 5 seconds and incubated for another 8 minutes at 55°C. RNase A was then added and the Eppendorf tube was incubated at room temperature for 15 minutes. 500 µL of Lysis buffer type 4 was added followed by another 10 minute incubation at room temperature. After a 5 second centrifugation step at 1,000 x g, the supernatant was transferred to a Bacteria Mini Column that was placed inside a collection tube and centrifuged for 1 minute at 11,000 x g. The resulting flowthrough was discarded and 500 µL of Lysis buffer type 4 was added. Following another 1 minute centrifugation step at 11,000 x g, the flowthrough was discarded and Wash buffer type 6 was added. The tube was then centrifuged for 3 minutes at 16,000 x g. The Bacteria Mini Column was the transferred to a new DNase-free eppendorf tube. 200 µL of pre-heated Elution buffer type 5 was added and the tube was incubated for 1 minute at room temperature. After a 1 minute spin at 11,000 x g, the flow through was retained, the genomic DNA was then stored at 4°C.

2.3 DNA Cloning

2.3.1 Restriction enzyme (RE) digestion

Single digest reactions involving only one restriction enzyme were typically performed in 20 μL volumes. This final volume would reach up to 60 μL when the digestion of a large quantity of DNA was required. The reaction mixture consisted of plasmid DNA, 10x RE buffer added accordingly to give a final concentration of 1x and 0.5-2 U of restriction enzyme. Sterile MilliQ water was added to the reaction mixture to reach the final volume required. The mixture was vortexed and incubated at 37˚C for 1-3 h.

Double digests involving two restriction enzymes were performed simultaneously when the buffering conditions required by both enzymes were compatible. The reaction mixture and conditions used were the same as those for single digest reactions. In some cases, the concentration of buffer varied depending on the requirements of the enzymes and the reactions were typically performed in 30 μL volumes.

Most of the restriction enzymes were heat inactivated as specified by Fermentas or New England Biolabs (NEB). In cases where the entire reaction was to be immediately loaded onto an agarose gel, heat inactivation was not carried out.

2.3.2 DNA amplification

Amplification of specific genes was performed via Polymerase Chain Reaction (PCR). Before carrying out a PCR reaction, primers were designed following a number of specifications. Custom designed primers were ordered from Sigma-Proligo and diluted in nuclease free water to reach a final concentration of 50 μM. The PCR primers for this study are listed in Table 2.4. The

Table 2.4: Primers used in this study

Primer Name	*Primer Sequence (5'-3')*	*Annealing site*	*Restriction site*
Cloning			
gtrIVNheIFnew	*TCAGCTAGCCTCGGTGGTGTGCAGCTC*	*Upstream of gtrIV gene*	*Nhe*I
gtrIVXbaIRnew	*TCATCTAGACCCCCCCAGGATAACTGTGGG*	*Downstream of gtrIV gene*	*Xba*I
phoF	*GTTCTGGAAAACCGGGCTGCTCAG*	*Beginning of phoA/lacZ coding sequence*	-
PCR-based Fusions for Topology			
V102HpaIRev	*TCAGTTAACCACGATTTTGGAAAGAGATGAC*	*In gtrIV gene*	*Hpa*I
Lp5Q146Rev	*CTGATAAAATGCTGTGAATTG*	*In gtrIV gene*	-
V155HpaIRev	*TCAGTTAACCGGAAGTAAAGTGATTACTATTTG*	*In gtrIV gene*	*Hpa*I
D169HpaIRev	*TCAGTTAACATCAAACCTTGATATTTCTTTTTTAAG*	*In gtrIV gene*	*Hpa*I
Lp10D407Rev	*GTCTAAATCTCTATATCCTTC*	*In gtrIV gene*	-
GtrIVCtermHpaIR	*TCAGTTAACCTTATAAATTCCTGATGCTACC*	*In gtrIV gene*	*Hpa*I
Sandwich fusion primers			
GtrIVR93NruIF	*CACATTTGATCTT**TCGCGA**TGGTCATCTCTTTCC*	*In gtrIV gene*	*Nru*I
GtrIVR93NruIR	*GGAAAGAGATGACCA**TCGCGA**AAGATCAAATGTG*	*In gtrIV gene*	*Nru*I
GtrIVK117NruIF	*GTAGAAATAAACAA**TCGCGA**ATATTCTTATCTTTTATTTC*	*In gtrIV gene*	*Nru*I
GtrIVK117NruIR	*GAAATAAAAGATAAGAATAT**TCGCGA**TTGTTTATTTCTAC*	*In gtrIV gene*	*Nru*I
GtrIVI136NruIF	*GCTATTTACTCCT**TCGCGA**CTTTCACAATC*	*In gtrIV gene*	*Nru*I
GtrIVI136NruIR	*GAATTGTGAAAG**TCGCGA**AGGAGTAAATAGC*	*In gtrIV gene*	*Nru*I
GtrIVA181NruIF	*GCGCAGCGCTAA**TCGCGA**CATCAAAATCTC*	*In gtrIV gene*	*Nru*I
GtrIVA181NruIR	*GAGATTTTGATG**TCGCGA**TTAGCGCTGCGC*	*In gtrIV gene*	*Nru*I
GtrIVR375NruIF	*CACTGATTTTTAT**TCGCGA**AAAATATTCTGTTG*	*In gtrIV gene*	*Nru*I
GtrIVR375NruIR	*CAACAGAATATTTT**TCGCGA**ATAAAAATCAGTG*	*In gtrIV gene*	*Nru*I
GtrIV Loop Deletions			
IVLp2DelSmaIF	*TCACCCGGGTATAAAAGCTCATTCAGTTATATTC*	*In gtrIV gene*	*Sma*I
IVLp2DelSmaIR	*TCACCCGGGATTATTCATCATCCATATGTC*	*In gtrIV gene*	*Sma*I
Lp8DelSmaIF	*TCACCCGGGCAAAAAATAAAAGACTCGC*	*In gtrIV gene*	*Sma*I
Lp8DelSmaIR	*TCACCCGGGATTGCTTTTATTGTATAGC*	*In gtrIV gene*	*Sma*I
Lp8P_Del1SmaR	*TCACCCGGGAGGAACATGGAAGGTTGCTTTAG*	*In gtrIV gene*	*Sma*I
Lp8P_Del2SmaIF	*TCACCCGGGCTTTTGATGCTAAAGCAACC*	*In gtrIV gene*	*Sma*I
Lp8P_Del2SmaIR	*TCACCCGGGTTTATTTCCCCATGCATCAAC*	*In gtrIV gene*	*Sma*I
Lp8P_Del3SmaIF	*TCACCCGGGTTTGATTTAGATAAGGGAGC*	*In gtrIV gene*	*Sma*I
IVLp6P.Del3FD1F	*TCACCCGGGGCTTTTATTGTATAGCAAAGTC*	*In gtrIV gene*	*Sma*I
IVLp6P.Del3FD2F	*TCACCCGGGGACGAATGCATTGGTGTTGATG*	*In gtrIV gene*	*Sma*I
IVLp6P.Del3FD2R	*TCACCCGGGATTTTCTTTTCCATTGCTTTTATTG*	*In gtrIV gene*	*Sma*I
IVLp6P.Del3FD3F	*TCACCCGGGGTTGATGCATGGGGAAATAAATTTG*	*In gtrIV gene*	*Sma*I
IVLp6P.Del3FD3R	*TCACCCGGGCACGGCCCACGATGGATTTTC*	*In gtrIV gene*	*Sma*I
IVLp6P.Del3FD4R	*TCACCCGGGACCAATGCATTCGTCCACGGC*	*In gtrIV gene*	*Sma*I

Table 2.4: Primers used in this study

Primer Name	*Primer Sequence (5'-3')*	*Annealing site*	*Restriction site*
Cloning			
gtrIVNheIFnew	*TCAGCTAGCCTCGGTGGTGTGCAGCTC*	*Upstream of gtrIV gene*	*Nhe*I
gtrIVXbaIRnew	*TCATCTAGACCCCCCAGGATAACTGTGGG*	*Downstream of gtrIV gene*	*Xba*I
phoF	*GTTCTGGAAAACCGGGCTGCTCAG*	*Beginning of phoA/lacZ coding sequence*	-
PCR-based Fusions for Topology			
V102HpaIRev	*TCAGTTAACCACGATTTTGGAAAGAGATGAC*	*In gtrIV gene*	*Hpa*I
Lp5Q146Rev	*CTGATAAAATGCTGTGAATTG*	*In gtrIV gene*	-
V155HpaIRev	*TCAGTTAACCGGAAGTAAAGTGATTACTATTTG*	*In gtrIV gene*	*Hpa*I
D169HpaIRev	*TCAGTTAACATCAAACCTTGATATTTCTTTTTTAAG*	*In gtrIV gene*	*Hpa*I
Lp10D407Rev	*GTCTAAATCTCTATATCCTTC*	*In gtrIV gene*	-
GtrIVCtermHpaIR	*TCAGTTAACCTTATAAATTCCTGATGCTACC*	*In gtrIV gene*	*Hpa*I
Sandwich fusion primers			
GtrIVR93NruIF	*CACATTTGATCTT**TCGCGA**TGGTCATCTCTTTCC*	*In gtrIV gene*	*Nru*I
GtrIVR93NruIR	*GGAAAGAGATGACCA**TCGCGA**AAGATCAAATGTG*	*In gtrIV gene*	*Nru*I
GtrIVK117NruIF	*GTAGAAATAAACAA**TCGCGA**ATATTCTTATCTTTTATTTC*	*In gtrIV gene*	*Nru*I
GtrIVK117NruIR	*GAAATAAAAGATAAGAATAT**TCGCGA**TTGTTTATTTCTAC*	*In gtrIV gene*	*Nru*I
GtrIVI136NruIF	*GCTATTTACTCCT**TCGCGA**CTTTCACAATC*	*In gtrIV gene*	*Nru*I
GtrIVI136NruIR	*GAATTGTGAAAG**TCGCGA**AGGAGTAAATAGC*	*In gtrIV gene*	*Nru*I
GtrIVA181NruIF	*GCGCAGCGCTAA**TCGCGA**CATCAAAATCTC*	*In gtrIV gene*	*Nru*I
GtrIVA181NruIR	*GAGATTTTGATG**TCGCGA**TTAGCGCTGCGC*	*In gtrIV gene*	*Nru*I
GtrIVR375NruIF	*CACTGATTTTTAT**TCGCGA**AAAATATTCTGTTG*	*In gtrIV gene*	*Nru*I
GtrIVR375NruIR	*CAACAGAATATTTT**TCGCGA**ATAAAAATCAGTG*	*In gtrIV gene*	*Nru*I
GtrIV Loop Deletions			
IVLp2DelSmaIF	*TCACCCGGGTATAAAAGCTCATTCAGTTATATTC*	*In gtrIV gene*	*Sma*I
IVLp2DelSmaIR	*TCACCCGGGATTATTCATCATCCATATGTC*	*In gtrIV gene*	*Sma*I
Lp8DelSmaIF	*TCACCCGGGCAAAAAATAAAAGACTCGC*	*In gtrIV gene*	*Sma*I
Lp8DelSmaIR	*TCACCCGGGATTGCTTTTATTGTATAGC*	*In gtrIV gene*	*Sma*I
Lp8P_Del1SmaR	*TCACCCGGGAGGAACATGGAAGGTTGCTTTAG*	*In gtrIV gene*	*Sma*I
Lp8P_Del2SmaIF	*TCACCCGGGCTTTTGATGCTAAAGCAACC*	*In gtrIV gene*	*Sma*I
Lp8P_Del2SmaIR	*TCACCCGGGTTTATTTCCCCATGCATCAAC*	*In gtrIV gene*	*Sma*I
Lp8P_Del3SmaIF	*TCACCCGGGTTTGATTTAGATAAGGGAGC*	*In gtrIV gene*	*Sma*I
IVLp6P.Del3FD1F	*TCACCCGGGGCTTTTATTGTATAGCAAAGTC*	*In gtrIV gene*	*Sma*I
IVLp6P.Del3FD2F	*TCACCCGGGGACGAATGCATTGGTGTTGATG*	*In gtrIV gene*	*Sma*I
IVLp6P.Del3FD2R	*TCACCCGGGATTTTCTTTTCCATTGCTTTTATTG*	*In gtrIV gene*	*Sma*I
IVLp6P.Del3FD3F	*TCACCCGGGGTTGATGCATGGGGAAATAAATTTG*	*In gtrIV gene*	*Sma*I
IVLp6P.Del3FD3R	*TCACCCGGGCACGGCCCACGATGGATTTTC*	*In gtrIV gene*	*Sma*I
IVLp6P.Del3FD4R	*TCACCCGGGACCAATGCATTCGTCCACGGC*	*In gtrIV gene*	*Sma*I

Chimera Studies			
GtrIVLp2BglIIF	*TCAAGATCTAATGGCGACTTCGATAGGGC*	*In gtrIV gene*	*Bgl*II
GtrIVLp2MluIR	*TCAACGCGTATACTCATAGTTCACAATTGAAC*	*In gtrIV gene*	*Mlu*I
GtrIVMluIF	*TCAACGCGTTCATTGAGTTATATTCTTTATTTATACG*	*In gtrIV gene*	*Mlu*I
GtrIVBglIIR	*TCAAGATCTCATCCATATGTAAATTGACTCC*	*In gtrIV gene*	*Bgl*II
GtrIcLp2BglIIF	*TCAAGATCTCACCATGGAAGCTGGTCAGG*	*In gtrIc gene*	*Bgl*II
GtrIcLp2MluIR	*TCAACGCGTATTGAAATCGTGACTTATATAAAATATAG*	*In gtrIc gene*	*Mlu*I
GtrIcBodyMluIF	*TCAACGCGTGTTACCGTTAAGTTATTAGCG*	*In gtrIc gene*	*Mlu*I
GtrIcVectorBglIIR	*TCAAGATCTAAAGACTGGCATCCACGATA*	*In gtrIc gene*	*Bgl*II
GtrIVLp8NruIF	*TCATCGCGAAAAATAACAGCTATCACTCC*	*In gtrIV gene*	*Nru*I
GtrIVLp8NruIR	*TCATCGCGATTTTTTGAACTTCATTCAAAAAAC*	*In gtrIV gene*	*Nru*I
GtrIVBodySmaIF	*TCACCCGGGGACTCTGCTCTATCAACTGTG*	*In gtrIV gene*	*Sma*I
GtrIVBodySmaIR	*TCACCCGGGGCTTGGCGAGAAAAATGAAAATG*	*In gtrIV gene*	*Sma*I
GtrIcLp10SmaIF	*TCACCCGGGGTACCTAAAAAAAGTAGCCAC*	*In gtrIc gene*	*Sma*I
GtrIcLp10SmaIR	*TCACCCGGGCCTGTTTGAACCTATAAGTTG*	*In gtrIc gene*	*Sma*I
GtrIcBodySmaIF	*TCACCCGGGAGTAATCGCACTTGTCATCAAAC*	*In gtrIc gene*	*Sma*I
GtrIcBodySmaIR	*TCACCCGGGGTCGTATTAATATCTTTCATAGTA*	*In gtrIc gene*	*Sma*I
Site-directed Mutagenesis			
D31ForA	*GATGAATAATGGC**GCA**TTCGATAGGGCCATAAC*	*In gtrIV gene*	-
D31RevA	*GTTATGGCCCTATCGAA**TGC**GCCATTATTCATC*	*In gtrIV gene*	-
D33ForwardA	*GAATAATGGCGACTTC**GCA**AGGGCCATAACACC*	*In gtrIV gene*	-
D33ReverseA	*GGTGTTATGGCCCT**TGC**GAAGTCGCCATTATTC*	*In gtrIV gene*	-
D31D33ForAA	*GATGAATAATGGC**GCA**TTC**GCA**AGGGCCATAAC*	*In gtrIV gene*	-
D31D33RevAA	*GTTATGGCCCT**TCG**GAA**TGC**GCCATTATTCATC*	*In gtrIV gene*	-
D49AFor	*CAGTTCTCACATGCT**GGC**ACATTACTTTATAC*	*In gtrIV gene*	-
D49ARev	*GTATAAAGTAATGT**GCC**AGCATGTGAGAACTG*	*In gtrIV gene*	-
E58AFor	*CTTTATACCCTGAAG**GCA**AATTTCAGTTCAATT*	*In gtrIV gene*	-
E58ARev	*AATTGAACTGAAATT**TGC**CTTCAGGGTATAAAG*	*In gtrIV gene*	-
E67AFor	*TCAATTGTGAACTAT**GCG**TATAAAAGCTCATTC*	*In gtrIV gene*	-
E67ARev	*GAATGAGCTTTTATA**CGC**ATAGTTCACAATTGA*	*In gtrIV gene*	-
E254AFor	*AAAAGCAATGGAAAA**GCA**AATCCATCGTGGGCC*	*In gtrIV gene*	-
E254ARev	*GGCCCACGATGGATT**TGC**TTTTCCATTGCTTTT*	*In gtrIV gene*	-
DE261AAFor	*CATCGTGGGCCGTG**GCCGCA**TGCATTGGTGTTG*	*In gtrIV gene*	-
DE261AARev	*CAACACCAATGCA**TGCGGC**CACGGCCCACGATG*	*In gtrIV gene*	-
GtrIVA261DFor	*AATCCATCGTGGGCCGTG**GAC**GCATGCATTGGTGTT*	*In gtrIV gene*	-
GtrIVA261DRev	*AACACCAATGCA**TGC**GTCCACGGCCCACGATGGATT*	*In gtrIV gene*	-
GtrIVA262EFor	*TCGTGGGCCGTGGCT**GAA**TGCATTGGTGTTGCT*	*In gtrIV gene*	-
GtrIVA262ERev	*AGCAACACCAATGCA**TTC**AGCCACGGCCCACGA*	*In gtrIV gene*	-
C263AFor	*TGGGCCGTGGACGAA**GCC**ATTGGTGTTGATGCA*	*In gtrIV gene*	-
C263ARev	*TGCATCAACACCAAT**GGC**TTCGTCCACGGCCCA*	*In gtrIV gene*	-
I264AFor	*GCCGTGGACGAATGC**GCT**GGTGTTGATGCATGG*	*In gtrIV gene*	-

I264ARev	*CCATGCATCAACACC**AGC**GCATTCGTCCACGGC*	*In gtrIV gene*	-
G265AFor	*GTGGACGAATGCATT**GCT**GTTGATCGATCCGGA*	*In gtrIV gene*	-
G265ARev	*TCCCCATGCATCAAC**AGC**AATGCATTCGTCCAC*	*In gtrIV gene*	-
I266AFor	*GACGAATGCATTGGT**GCT**GATGCATGGGGAAAT*	*In gtrIV gene*	-
I266ARev	*ATTTCCCCATGCATC**AGC**ACCAATGCATTCGTC*	*In gtrIV gene*	-
D267AFor	*GAATGCATTGGTGTT**GCT**GCATGGGGAAATAAA*	*In gtrIV gene*	-
D267ARev	*TTTATTTCCCCATGC**AGC**AACACCAATGCATTC*	*In gtrIV gene*	-
W269AFor	*ATTGGTGTTGATGCA**GCG**GGAAATAAATTTGAT*	*In gtrIV gene*	-
W269ARev	*ATCAAATTTATTTCC**CGC**TGCATCAACACCAAT*	*In gtrIV gene*	-
D276ForwardA	*GGAAATAAATTTGATTTA**GCT**AAGGGAGCTGTTACC*	*In gtrIV gene*	-
D276ReverseA	*GGTAACAGCTCCCTT**AGC**TAAATCAAATTTATTTCC*	*In gtrIV gene*	-
ED283AAFor	*GCTGTTACCACA**GCAGCT**GCTGGAGCCTGC*	*In gtrIV gene*	-
ED283AARev	*GCAGGCTCCAGC**AGCTGC**TGTGGTAACAGC*	*In gtrIV gene*	-
E294AFor	*TTCGCAAAAAATAAA**GCA**AGGGGATTGAAGGAT*	*In gtrIV gene*	-
E294ARev	*ATCCTTCAATCCCCT**TGC**TTTATTTTTTGCGAA*	*In gtrIV gene*	-
D317For	*CTCCTTCTTCCTTTT**GCT**TCTGGAGTTCGAACC*	*In gtrIV gene*	-
D317Rev	*GGTTCGAACTCCAGAA**AGC**AAAAGGAAGAAGGAG*	*In gtrIV gene*	-
ED326AAFor	*GAACCCAATTAACT**GCAGCT**TACTTCCATGTC*	*In gtrIV gene*	-
ED326AARev	*GACATGGAAGTAA**GCTGC**AGTTAATTGGGTTC*	*In gtrIV gene*	-
E334AFor	*CTTCCATGTCTTCAA**GCC**AAATAAGCTAATAATAAG*	*In gtrIV gene*	-
E334ARev	*CTTATTATTAGCTTATTT**GGC**TTGAAGACATGGAAG*	*In gtrIV gene*	-
E347AFor	*AGTGTTTTTTTGAAT**GCA**GTTCAAAAAATAAAA*	*In gtrIV gene*	-
E347ARev	*TTTTATTTTTTGAAG**AGC**ATTCAAAAAAACACT*	*In gtrIV gene*	-
Sequencing			
PhoSeqNewR	*GCACCTTCGGCATAATTACGTGC*	*173 bp into phoA/lacZ coding sequence*	
M13R	*GGAAACAGCTATGACCATG*	*pBS KS multiple cloning site*	

Primers have been grouped into categories that correspond to their main use. Some primers were used in multiple studies. Restriction enzyme sites on the primers are underlined. For primers used in mutagenesis, the bases that have been changed are in bold italics.

design of the oligionucleotide sequence for a pair of forward and reverse primers was based on a matching melting temperature (T_m), which was calculated using the formula below:

T_m (°C) = (2 x number of A + T bases) + (4 x number of G + C bases)

A typical PCR reaction (3x) consisted of a mastermix of 103.5 µL of sterile MilliQ water, 3 µL of each 2.5 µM primer stock (final concentration: 0.05 µM), 3 µL of 10 mM deoxynucleotide triphosphate (dNTP) mix (final concentration: 0.2 mM), and 15 µL of 10x *Pfu* ultraII polymerase buffer (Stratagene; to give 1x) are added together into an 1.5mL microcentrifuge tube. From this mastermix, 42-48 µL was aliquoted out into a PCR tube. 2-8 µL of DNA template (diluted 1/20) was added to give a final mastermix volume of 50 µL. Finally, 0.2 µL of *Pfu* ultraII poymerase (Stratagene) was added to each PCR tube. Table 2.5 lists the different protocols used for the respective PCR reactions. A T3 thermal cycler (Biometra) was used. The primer annealing temperature was typically 3-5°C below the melting temperature (T_m) of the primer (excluding non-complementary overhangs). The extension step used was 1 minute per kb of DNA. For the amplification of entire plasmids, the primer annealing and extension temperatures were 55°C and 68°C, respectively, with 25 cycles amplification. Upon completion of the PCR, the holding temperature was 4°C.

2.3.2.1 Colony PCR

Colony PCR was used to amplify *gtrIV* from the wildtype serotype 4 strains. The protocol was adapted from Schuch and Maurelli (1997). Colonies were smeared into 1.5 mL microcentrifuge tubes, resuspended in 25 µL of 0.5 M NaOH and let at RT for 30 min. 25 µL of 1 M Tris (pH 8) and 450 µL of sterile MilliQ water were then added to each microcentrifuge tube. The PCR mastermix had the same composition as described above with *Pfu* ultraII

Table 2.5: Typical PCR conditions used in this study.

Initial Denaturation	Denaturation	Primer annealing	Extension	Final extension
PCR protocol for cloning and colony PCR				
95°C, 2 min 1x	95°C, 30 s 30x	3°C below lowest primer T_m, 30 s 30x	72°C, 1 min per kb 30x	72°C, 7 min 1x
Whole plasmid amplification and site-directed mutagenesis				
95°C, 30 s 1x	95°C, 30 s 25x	55°C, 1 min 25x	68°C, 1 min per kb 25x	-

polymerase. The total volume depended on the number of reactions. The reaction mixes consisted of 45 μL PCR mastermix and 5 μL of lysed colony to make a total volume of 50 μL.

2.3.3 Removal of template DNA through DpnI treatment

To remove methylated template DNA from PCR products, 10 U of *Dpn*I (NEB) was added directly to a completed PCR reaction. The tube was incubated at 37°C for 1-2 h. In some cases, the digest was left overnight in the 37°C waterbath.

2.3.4 DNA Quantification

The concentration and purity of DNA was estimated using the Nanodrop ND-1000 Spectrophotometer by Nanodrop Technologies. A 2 μL aliquot of each sample was used for analysis.

2.3.5 DNA Ligations

Both the vector and insert DNA were digested with two restriction enzymes. This forms cohesive overhangs. A vector to insert ratio of 3:1 was used to help the insert ligate to the vector. Typically, ligation reactions were made up to 20 μL consisting of 6 μL digested vector, 1-2 μL insert DNA, sterile MilliQ water, 1x ligase buffer (Promega) and 1.2 U of T4 DNA ligase (Promega). Sticky end ligation reactions were incubated overnight at 4°C while blunt-end ligation reactions were incubated overnight at 15°C. All ligations were stored at -20°C. For transformation into rubidium chloride competent cells, the entire ligation mix was used. For transformation into electrocompetent cells, smaller volumes (1-3 μL) were used.

2.3.6 DNA sequencing

For sequencing, DNA was prepared using the Axygen Axyprep Plasmid Miniprep Kit (section 2.2.2). The extracted DNA was then quantified by using the NanoDrop ND-1000 spectrophotometer (section 2.3.4). Approximately 150-200 ng of DNA, 1 μL of Terminator Ready Reaction Premix, 3.5 μL of 5x dilution buffer, 1.5 μL of 2.5 μM primer and sterile MilliQ water up to 20 μL were mixed together into a PCR tube. The Terminator Ready Reaction Premix and 5x dilution buffer were obtained from the Biomolecular Resources Facility (BRF), John Curtin School of Medical Research, Australian National University. The tube was vortexed, pulse centrifuged and then subjected to 25 cycles of the following conditions in the GenAmp thermal cycler: 96°C 10 s, 50°C 5 s, 60°C 4 min. The reaction was held at 4°C.

The 20 μL reaction was transferred into a microcentrifuge tube and the following were added: 2 μL of 3 M sodium acetate, 2 μL of 125 mM EDTA, and 50 μL of 100% ethanol. The mixture was vortexed and centrifuged for 30 min. The DNA pellet was washed with 250 μL of 70% ethanol followed by a 2 minutes centrifugation. This was repeated 2 times. After the three washes, the pellet was then dried under vacuum for 5 minutes (Savant SpeedVac SC100; high drying rate). After the tube was dry, it was stored away from light at 4°C until it was delivered to the BRF to be sequenced with ABI 3730 capillary sequence analyser (PE Applied Biosystems). The resulting sequencing data was analysed using a bioinformatics program, pDraw32 (AcaClone Software, http://www.acaclone.com) and ClustalW (European Bioinformatics Institute, Larkin *et al.*, 2007).

2.4 Agarose Gel Electrophoresis

Agarose gel electrophoresis was carried out using 0.7% agarose gels that contained ethidium bromide (10 μg/mL) in 0.5x Tris-Borate-EDTA (TBE) buffer (Appendix A). The

running buffer, which consisted of 0.5x TBE buffer, was poured into the gel tank before the samples were loaded. Approximately 0.2 volumes of blue loading dye (Appendix A) was added to each sample prior to loading. The voltage was usually set between 70-90 V and the gel was run until the blue loading dye moved to the end of the gel. The size marker used was *Eco*RI-digested SPP-1 bacteriophage DNA (GeneWorks) and is shown in Figure 2.2. In order to visualise the DNA bands, the gel was viewed under ultraviolet light by utilising a CCD camera and NIH image software (Mitsubishi).

2.4.1 Purification of DNA from agarose gels

By using a sterile scalpel, the DNA bands were excised from the agarose gel. The gel slice was then placed into 1.5 mL microcentrifuge tubes and weighed. The DNA was extracted using the Wizard SV Gel and PCR Clean-up System (Promega). 10 µL of Membrane Binding Solution was added to per 10 mg of gel slice. The tubes were then vortexed and incubated in a 65°C waterbath for 15 minutes. An SV column was then added to a collection tube and the dissolved gel mixture was then transferred to this assembly and left at room temperature for 1 minute. The tubes were then centrifuged for 1 minute. The flowthrough was discarded and the SV column was reinserted into the collection tube. 700 µL of Membrane Wash Solution was added to the column assembly and centrifuged for 1 minute. The resulting flowthrough was discarded. This washing step was repeated again with 500 µL of Membrane Wash Solution, followed by a 5 minute centrifugation. The collection tube was emptied and the column assembly was re-centrifuged for 1 minute with the microcentrifuge lid open to allow evaporation of any residual ethanol. The minicolumn was then transferred to a clean 1.5 mL microcentrifuge tube. 40µL of Nuclease-Free water was added to the column. Following a 1 minute incubation at room temperature, the tubes were centrifuged for 1 minute. The minicolumn was discarded and the DNA was stored at -20°C.

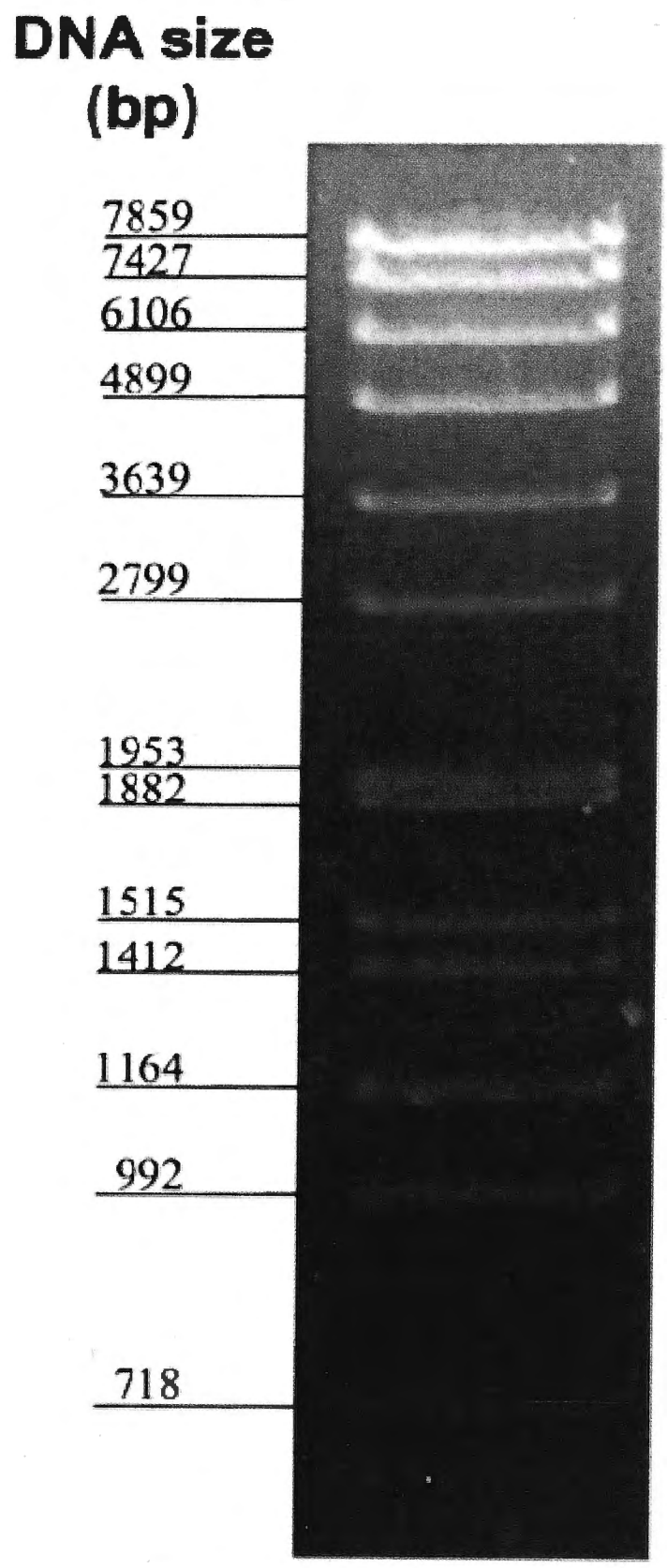


Figure 2.2: The *EcoRI* digest of the *Bacillus subtilis* SPP-1 bacteriophage genome. The digest was run in parallel with the DNA samples as a DNA size marker (adapted from GeneWorks).

2.5 Bacterial Transformation

2.5.1 Preparation of electrocompetent cells

The protocol for the preparation of electrocompetent cells is from Dower *et al.* (1988). All non-centrifugal steps to prepare the electrocompetent cells were carried out in a cold room at 4°C. XL1-Blue and JM109 were grown overnight in 20 mL of LB broth. SFL1616 was grown overnight in 20 mL of LB broth supplemented with Cm and Kan. The 20 mL overnight cultures were then diluted (1/100) into 1 L of LB broth that was pre-warmed at 37°C. The diluted cultures were then incubated at 37°C while being vigorously shaken at 230 rpm. The cultures were continuously incubated for 2-4 hours until the cultures reached log phase, which was measured by a Cary50 Bio Spectrophotometer (Varian) when the culture had an A_{600} of 0.6-0.8. The 1 L cultures were then chilled on ice for 30 minutes before they were aliquoted into duplicate 500 mL centrifuge bottles (Sorvall) that were pre-chilled at 4°C. The bacteria were then formed into a pellet by centrifugation at 3,000 x g at 4°C for 10 minutes (SLA-3000; Sorvall RC-5C Plus Superspeed Centrifuge). After centrifugation, the supernatant was discarded and the bacterial pellet was washed by resuspending it in 200 mL of ice-cold MilliQ water followed by centrifugation as before. The washing step was repeated twice more. The pellets were twice resuspended in 50 mL of ice-cold 10% glycerol (v/v) and centrifuged as before for the final time. The supernatant was discarded leaving behind a residual glycerol solution of about 1 mL for the bacterial pellet to be re-suspended in. 50 µL aliquots were transferred into sterile 1.5 mL Eppendorf tubes that were pre-chilled at 4°C. The bacterial aliquots were then stored at -80°C until they were required for transformation.

2.5.2 Transformation into electrocompetent cells

Electrocompetent cells were taken from the -80°C freezer and thawed on ice. About 1-3 μL of DNA was mixed with the cell suspension. The suspension was transferred into a chilled Bio-Rad electroporation cuvette and tapped to the bottom prior to electroporation. Using a BioRad Genepulser, a pulse of 2.5 kV was supplied. 1 mL of LB broth was then added to the cells in order to recover them. After mixing with the pipette, the mixture was transferred into a microcentrifuge tube. The cells were incubated at 37°C with shaking for 1 h. The tube was centrifuged for 1 minute and most of the supernatant solution was discarded, leaving behind about 100 μL of supernatant solution. The cells were resuspended in the remaining supernatant and spread onto an LB agar plate containing the appropriate antibiotic.

2.5.3 Determination of transformation efficiency

By determining the number of colonies obtained by transformation of a known concentration of pBC SK, the efficiency of competent cells was calculated. 1/10, 1/100, 1/1000 and 1/10,000 dilutions were performed to obtain a plate with an appropriate number of colonies. The efficiency was calculated using the following formula:

Number of transformed cells (N) = Number of colonies x Dilution factor

N/μg DNA (efficiency) = N/amount of DNA transformed (μg)

2.5.4 Selecting Transformed Colonies

Transformed bacteria were selected for their ability to grow on antibiotic-containing LB agar plates. The plasmids used in this study contained ampicillin, chloramphenicol or kanamycin resistance genes. The concentrations used for selection were stated in section 2.1.

2.6 Bioinformatics

Six topology prediction programs that were available on the Internet were used to examine the GtrIV protein sequence for the presence of hydrophobic regions. They were HMMTOP (www.enzim.hu/hmmtop/; Tusnady & Simon, 1998), SOSUI (http://sosui.proteome.bio.tuat.ac.jp/sosuiframe0.html; Hirokawa *et al.*, 1998), TMpred (www.isrec.isbsib.ch/software/TMPRED_form.html; Hofmann & Stoffel, 1993), DAS (www.sbc.su.se/~miklos/DAS/; Cserzo *et al.*, 1997), TopPred (www.sbc.su.se/~erikw/toppred2/; von Heijne, 1992) and finally, TMHMM (www.cbs.dtu.dk/services/TMHMM1.0/; Sonnhammer *et al.*, 1998). In addition, to compare the sequence and amino acid homology between GtrIV and the other wildtype sequences, the web-based program ClustalW (European Bioinformatics Institute, Larkin *et al.*, 2007) was used.

2.7 Alkaline Phosphatase and β-Galactosidase Assays

Alkaline phosphatase (AP) and β-galactosidase (BG) assays were carried out to quantify the AP and BG activities, respectively. The AP assay was first described by Manoil *et al.* (1990) while the BG assay was first described by Miller (1992). Both assays were performed in parallel. Two experiments were done simultaneously for each strain, with duplicates for each experiment. A negative control (B1752 carrying pNV1473), was included in each assay. This was done so that the background activities could be subtracted from the experimental activity values.

2.7.1 Alkaline phosphatase assay

0.5 mL aliquots of overnight cultures (5 mL) were added to each of two 9.5 mL aliquots of LB broth (1/20 dilution) supplemented with the appropriate antibiotic. The cells were incubated at 37°C with shaking until they reached log phase. This took approximately 2-3 h. The optimal OD_{600} was 0.6. 1 mM IPTG was used to induce the samples. The induced samples were

incubated at 37°C with shaking for 1 h. A 1 mL aliquot of each culture was then centrifuged for 3 min. Cold 10 mM Tris (pH 8)/10 mM $MgSO_4$ was used to wash the pellets. The pellets were then resuspended in 1 mL of cold 1M Tris (pH 8). Upon resuspension, 0.1 mL aliquots of cells were diluted in 0.9 mL of cold 1 M Tris (pH 8). At this point, the OD_{600} was measured. The remaining steps were performed in duplicate. 0.1 mL of washed cells were added to 0.9 mL of 1 M Tris (pH 8)/0.1 mM $ZnCl_2$. The cells were then permeabilised by adding 50 µL of 0.1% SDS and 50 µL of chloroform. After vortexing the tubes, they were incubated at 37°C for 5 minutes and then placed on ice for another 5 min.

To begin the assay, 0.1 mL of 0.4% *p*-nitrophenyl phosphate dissolved in 1M Tris (pH 8) was added to each tube. The tubes were inverted and placed in a 37°C water bath. The time was recorded immediately after the tubes were placed into the water bath. The reaction was stopped when it had turned pale yellow in colour. This was done by adding 0.12 mL of 1:5 0.5 M EDTA (pH 8):1 M KH_2PO_4 and placing the tubes on ice. The time was noted at this point. For reactions in which no yellow colour developed, they were stopped after 2 hours. The OD_{420} and OD_{550} were measured. These two wavelengths were chosen because the OD_{420} measures a combination of absorbance by *o*-nitrophenol and light scattering by cell debris. On the other hand, the OD_{550} corrects for light scattering. An assay mixture without cells was used as a blank. The AP activity was calculated in units activity as follows:

$$\text{Units activity} = \frac{[OD_{420} - (1.75 \times OD_{550})]1000}{\text{Time (min)} \times OD_{600} \times \text{volume of cells (mL)}}$$

2.7.2 β-galactosidase assay

Cells were diluted, grown and induced with IPTG as described for the AP assay. After the induction step, 2 mL of cells were pelleted by centrifuging for 3 min, and the supernatant solution was discarded. 2 mL of cold Z buffer (Appendix A) was used to resuspend the cell pellet. The OD_{600} was measured using Z buffer as a blank. Subsequent steps were performed in duplicate. Cells were diluted 1/2 by adding 0.5 mL of cells to 0.5 mL of Z buffer to give a final volume of 1 mL. At this point, 100 μL of chloroform and 50 μL of 0.1% SDS were added. After vortexing, the tubes were incubated for 5 minutes at 28°C. 0.2 mL of substrate, *o*-nitrophenyl-β-D-galactoside (ONPG; 4 mg/mL dissolved in phosphate buffer (Appendix A)) was added to each tube to start the reaction. The time of addition was recorded and the tubes were vortexed and incubated at 28°C. When a yellow colour had developed, the reactions were stopped by the addition of 0.5 mL of 1 M Na_2CO_3. At this point the time was recorded again. As with the AP assay, reactions that did not develop any colour were stopped after 2 h. The tubes were vortexed and 1 mL was transferred to a microcentrifuge tube and centrifuged for 5 minutes to remove debris and chloroform. The supernatant solution was then transferred to new tubes. The OD_{420} and OD_{550} were measured for each sample. The same formula used to determine units of AP activity was used to determine the units of BG activity.

2.7.3 Calcuating the Normalised Activity Ratios (NAR)

After obtaining the AP and BG activities for each fusion in the dataset, the normalised activity ratios were calculated as follows:

NAR = (AP/highest AP) ÷ (BG/highest BG)

2.8 Site-Directed Mutagenesis

Site-directed mutagenesis was performed as described in the protocol by Stratagene (QuikChange Site-Directed Mutagenesis Kit) (Figure 2.3). The following parameters were used when designing mutagenic primers:

1. Primers were designed to be about 25-35 bases long
2. The desired mutations were in the middle of the primer
3. C or G bases were on the 3' and 5'ends of the primers
4. Both forward and reverse primers that carry the same mutation were designed to anneal to the same sequence but on opposite strands.

Upon arrival of mutagenic primers, PCR amplification was carried out (section 2.3.2). A small aliquot (5 μL out of a 50 μL reaction) of the PCR product was then run on an agarose gel (section 2.4) to check if amplification had worked. The PCR product was then treated with *Dpn*I (section 2.3.3) and transformed into electrocompetent XL1-Blue cells using the protocol described in section 2.5.2. Transformant colonies were patched onto a masterplate and about 3-4 colonies were inoculated overnight so that the plasmid DNA could be isolated (section 2.2.2) and sequenced (section 2.3.6) to check the desired mutation.

2.9 Slide agglutination assays

The function of Gtr proteins was tested by transformation into a serotype Y *S. flexneri* strain containing *gtrA* and *gtrB* (SFL1616). Serotype conversion indicated a functional Gtr protein. Therefore, 3-4 colonies were streaked onto LB agar plates containing the appropriate antibiotics. Once grown, these lawns were used in slide agglutination tests. Cells were mixed directly into a drop of *S. flexneri* specific antisera on a glass slide. The slide was then rocked gently back and forth and the mixture was observed for agglutination. Agglutinations that

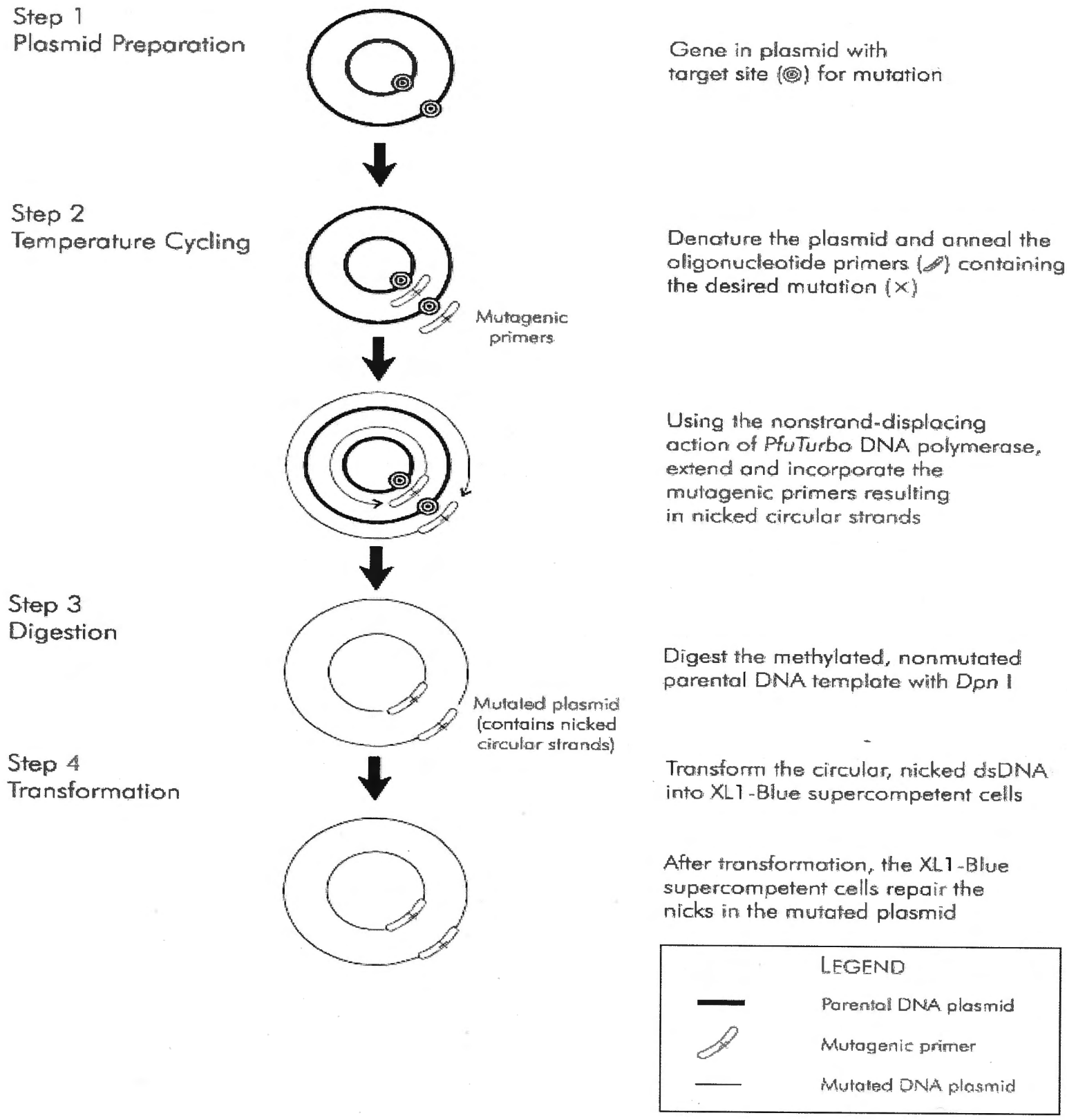


Figure 2.3: A summary of the steps involved in site-directed mutagenesis. Image taken from Stratagene manual.

occurred within one minute were considered positive. For a negative control, cells were mixed in phosphate buffered saline (PBS) instead of *S. flexneri* specific antisera.

2.10 Membrane protein preparation

The membrane protein isolation method is a modified version of that described by Morona *et al.* (1995). XL1-Blue cells were grown in LB to mid-log (OD_{600} of 0.6) and pelleted using a Sorvall SLA1500 rotor (3,000 x g, 10 min, 4°C) and resuspended in 1 mL of 20% (w/v) sucrose, 30 mM Tris-HCl pH 8.1, transferred to SS-34 tubes and chilled on ice. 0.1 mL of 1 mg/mL lysozyme in 0.1M EDTA pH 7.3 was added to the cells for 30 minutes on ice. The cells were collected again as described above using a Sorvall SS-34 rotor and the pellet frozen for 30 minutes at -80°C. The pellet was thawed and resuspended vigorously in 6 mL 3 mM EDTA, pH 7.3. The cells were completely lysed by passing them twice through a French Press at 15,000 psi. Unlysed cells and inclusion bodies were removed by slow centrifugation as described above (7,000 rpm, 10min, 4°C). Membrane proteins were sedimented by high speed centrifugation using a 50Ti or 80Ti rotor spun at 35, 000 rpm for 90 minutes at 4°C. The pellet was then resuspended in 200 µL sterile Milli Q H_2O.

2.10.1 Protein quantification by BCA assay

To calculate the protein concentration in the membrane protein preparation, a BCA protein assay (Thermo Scientific) was carried out. A 2.0 mg/mL stock of BSA was prepared. Nine diluted standards of BSA with the following concentrations were made using this stock solution (2.0 mg/mL, 1.5 mg/mL, 1.0 mg/mL, 0.75 mg/mL, 0.5 mg/mL, 0.25 mg/mL, 0.125 mg/mL, 0.025 mg/mL and 0 mg/mL). The BCA working reagent (WR) was also prepared by mixing 50 parts of Reagent A with 1 part of Reagent B. The total volume of WR required was calculated by the following formula:

Total Volume of WR required = (No. of Standards + No. of Unknowns) x (No. of Replicates) x (Volume of WR per Sample)

After the nine BSA standards and WR had been prepared, 25 µL of each standard and protein sample was loaded into a 96-well microplate. 200 µL of the WR was then added to each of the wells. The microplate was gently shaken and covered before being incubated at 37°C for 30 min. The plate was left to cool at room temperature before the absorbance levels of the standards and samples were measured on a plate reader at a wavelength of 562 nm (Gen5-Biotek Instruments). To determine the protein concentration of each unknown sample, a standard curve of Blank-corrected 562 nm measurement for each BCA standard versus mg/mL was created and used to determine the protein concentration of the unknown samples.

2.11 Bacterial lipopolysaccharide (LPS) preparation

This method was adapted from Hitchcock & Brown (1983) with several modifications. Overnight cultures were diluted 1/100 in LB containing the appropriate antibiotics and incubated for 2 - 3 hours at 37°C until OD_{600} of 0.6 was reached. 1.5 ml of culture was then spun down and the pellet was resuspended in 80µl sample loading buffer (4% SDS, 160mM Tris-HCl, 20% glycerol, 10% β-mercaptoethanol). Proteinase K was then added to each sample at a concentration of 50 mg/mL and the samples were incubated overnight at 56°C and stored at -20°C. Before samples are run on 12% SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE) gel, 0.5µl β-mercaptoethanol was added to the samples and boiled for 10 min.

2.12 Sodium dodecyl sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS page electrophoresis was carried out using a 4% stacking and 12% resolving SDS gel (Appendix A). SDS gels were set in 1 x SDS-PAGE buffer (Appendix A) before either LPS or membrane fraction samples were loaded on the gel. A Pierce Blue Prestained Protein Molecular Weight Marker (Thermo Scientific) or PageRuler™ Prestained Protein Ladder (Fermentas) was loaded in parallel with the samples as a protein size marker (Figure 2.4.). The gels were run by electrophoresis at 60-80V until the blue loading dye had run off the gel.

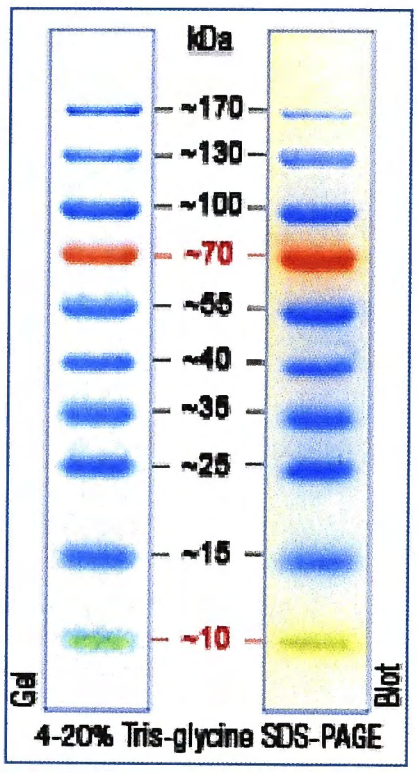
For SDS gel electrophoresis of membrane protein fractions, equal protein concentrations of each sample (estimated by a protein BCA assay) were aliquoted into sterile 1.5 mL Eppendorf tubes. 6x SDS blue loading dye was then added to each tube to obtain a final concentration of 1x SDS blue loading dye before the samples were loaded on the SDS gel.

Before SDS gel electrophoresis could be carried out on the bacterial LPS, 10 µL of each LPS preparation was aliquoted into a sterile 1.5 mL Eppendorf tube. 0.5 µL of 2-mercaptoethanol was added to each sample and the samples were boiled at 100°C for 5 min. The condensation was spun down by centrifugation for 5 seconds and each sample was loaded into the SDS gel.

2.13 Coomassie Staining of Membrane Protein Preparation

Following SDS gel electrophoresis of membrane protein samples (Section 2.12), the gel was then stained with 200 mL of Coomassie brilliant blue solution (Appendix A) overnight. The gel was destained with 100 mL of destaining solution for 15 minutes (Appendix A). The

a)



b)

Protein	MW
Myosin	215K
Phosphorylase B	120K
BSA	84K
Ovalbumin	60K
Carbonic Anhydrase	39.2K
Trypsin Inhibitor	28K
Lysozyme	18.3K

Figure 2.4: a) The PageRuler™ Prestained Protein Ladder is a 3-color ladder with 10 proteins covering a molecular weight range from 10 to 170 kDa (Fermentas). b) The appearance and molecular weight (MW) of constituent proteins in the Blue Prestained Protein Molecular Weight Marker Mix (Thermo Scientific).

destaining step was repeated twice. The gel was then sealed in a plastic sleeve containing 2 mL of 10% glycerol (v/v).

2.14 Silver staining of Bacterial LPS

10 μL of each LPS preparation was aliquoted into a sterile 1.5 mL Eppendorf tube. 0.5 μL of 2-Mercaptoethanol was added to each sample and the samples were boiled at 100°C for 5 min. The samples were spun down by centrifugation for 5 seconds before each sample was loaded into the SDS gel (Section 2.12). After SDS gel electrophoresis, the SDS gel was fixed overnight with 100 mL of fixing solution (Appendix A). The following day, the fixing solution was discarded and the gel was washed twice with 7.5% acetic acid (v/v) for 30 minutes each time. The gel was then soaked in oxidising solution (Appendix A) for 5 min. The oxidising solution was then discarded and the gel was washed with 100 mL MilliQ water for 30 min. The washing step was repeated thrice. The gel was then soaked in silver stain solution (Appendix A) for 10 minutes. The silver stain solution was discarded and the gel was washed three times with 100 mL of MilliQ water for 5 minutes each. After the washes, the gel was soaked in developer solution (Appendix A) for 5-10 minutes until the desired band intensity was obtained. To stop development, the gel was blocked using 1% acetic acid for 15 minutes. The gel had a final wash of 100 mL of MilliQ water before being stored in a plastic sleeve containing 2 mL of 10% glycerol (v/v).

2.15 Western Immunoblotting

Prior to carrying out the Western blotting, membrane protein and LPS samples were run on SDS page gels until the dye front ran out of the gel. 15 minutes prior to the transfer, one sheet of PVDF membrane (Millipore) was soaked in methanol and two sheets of Whatman paper and two support pads were soaked in transfer buffer (Appendix A). The transfer sandwich apparatus

was assembled in the following order; black face of the sandwich holder at the bottom, support pad, Whatman paper, SDS gel, PVDF membrane, Whatman paper, support pad, white face of the sandwich holder at the top. The sandwich apparatus was then closed and placed in the BioRad Mini Trans-Blot Transfer cell apparatus. The tank was then filled with 1x transfer buffer and the samples were transferred at a maximum current of 250mA at 4°C for 2 hours with gentle stirring.

After the transfer, the membrane was blocked overnight with 50 mL of blocking solution containing 5% skim milk powder in PBS (Appendix A). After being blocked overnight, the membrane was washed three times with 50 mL of 1x PBS + 0.05% Tween20 for 15 minutes each time (Appendix A). The membrane was then transferred to a plastic sleeve and 5 mL of the appropriate primary antibody diluted in 1% skim milk and 0.1% Tween20 was added to the plastic sleeve. The plastic sleeve was sealed and vigorously shaken at 4°C for 2 hours. After the 2 hours, the membrane washed three times with 50 mL 1 x PBS + 0.05% Tween20 for 10 minutes each time (Appendix A). The membrane was then transferred to a plastic sleeve and 5 mL of the appropriate secondary antibody mix diluted in 1% skim milk and 0.1% Tween20 was added. The sleeve was sealed and vigorously shaken at 4°C for 1 hour. Following this, the membrane was washed twice with 1x PBS + 0.05% Tween 20 and twice with 1x PBS for 10 minutes each. Finally, the membrane was then covered with 4 mL of SuperSignal West Pico Stable Peroxide Solution and SuperSignal West Pico KLuminol/Enhancer Soltuion (Thermo Scientific) in a ratio of 1:1. The tray was covered with foil and the membrane was left at room temperature for 5 minutes. The reactions were detected either by X-ray film (GE Healthcare) or viewed under the Fusion Chemiluminescence Camera (Fisher Biotech).

Primary antibodies used were either Mouse anti-alkaline phosphatase (Chemicon) diluted 1:1000 for membrane protein preparations, Type IV antisera (Denka Seiken) (specific to serotype 4a) diluted 1:100 or serotype Ic-specific MASFIc (monoclonal antibody *Shigella*

flexneri Ic) monoclonal antibody (Reagensia) diluted 1:500 for bacterial LPS extractions. The secondary antibodies used were Goat anti-mouse IgG horse radish peroxidase (HRP)-conjugated (Sigma) diluted 1:8000 for membrane protein preparations while Goat anti-rabbit HRP-conjugated Ig (Sigma) diluted 1:1000 and anti-mouse IgM peroxidise conjugate (Sigma) diluted 1:5000 were used for profiling bacterial LPS extraction.

2.16 Southern Hybridization

2.16.1 Transferring of DNA from gel to membrane

Genomic DNA digests were run on 0.7% agarose gel until the blue loading dye was run off the gel. Following this, the gel was photographed under ultraviolet light by utilising a CCD camera and NIH image software. The wells on the gel were cut off and the gel was soaked in 0.25M HCl (Appendix A) for 15 minutes at room temperature with gentle shaking. The gel was then rinsed with MilliQ water before being soaked in denaturing solution (Appendix A) for 30 minutes at room temperature with gentle shaking. The gel was rinsed again with MilliQ water. Following this, the gel was soaked with neutralising solution (Appendix A) for 15 minutes at room temperature with gentle shaking. After another rinse in MilliQ water, the gel placed in the transfer apparatus. Care was taken to ensure that the Hybond N+nucleic acid transfer membrane from GE was cut such that it was only slightly larger than the gel.

The following day, the gel was removed from the apparatus with the transfer membrane attached to it. The positions of the gel wells were marked on the membrane. Then, the membrane was washed 5x SSC for 5 minutes before being placed on top of aluminium foil and baked in vacuum oven at 80°C for 2 hours. After baking, the membrane was sandwiched between hybond cover papers and stored in the fridge until it was used.

2.16.2 Labelling of probe with Digoxigenin (DIG)

The probe was labelled by adding 100ng of DNA (3.8 kb *gtrIV* cassette fragment) to 16 µL of MilliQ water in a 1.5 mL Eppendorf tube. This mixture was boiled for 10 minutes to denature the DNA and then placed on ice for 1 minute. 4 µL of DIG-High Prime (Roche) was then added to the tube and left overnight (up to 20 hours) at 37°C. The next day, 1 µl of 0.5M EDTA (pH 8) was added to the tube to stop the labelling reaction. The tube was then incubated for 10 minutes at 65°C. The desired amount of probe was added to the hybridization buffer (20-50 ml) to commence the Southern hybridization process.

2.16.3 Hybridization of membranes

An appropriate volume of DIG Easy Hybrid (10 mL per membrane) was pre-heated to the hybridization temperature of 65°C. The membranes were then placed in hybridization tubes with the preheated DIG Easy Hybrid for 30 minutes – 2 hours in a Hybridization oven at 65°C. Meanwhile, the DIG-labelled probe was boiled to denature any remaining double strands. After hybridization, the freshly denatured probe was added and left overnight. The next day, the membrane was washed for 5 minutes with 50 mL of DIG Washing Buffer. The membrane was then treated with 100 mL of DIG 1x blocking buffer. Following this, the membrane was incubated with 10 mL α-DIG-Antibody-solution (Roche) for 30 minutes. Two 15 minute washes with 100 mL DIG Washing Buffer then ensued. After the washes, the membrane was equilibrated in 20 mL of DIG Detection Buffer. The membrane was incubated 5 minutes at 15-25°C with 10 mL of CSPD-Solution. Using a pair of tweezers, the membrane was placed (DNA side up) in an x-ray cassette or hybridization bag. After incubating for 10 minutes, the membrane was viewed under the Fusion Chemiluminescence Camera (Fisher Biotech).

Chapter 3

Elucidating the topology of GtrIV

3.1 Introduction

Determining the topologies of the Gtrs, or their dispositions relative to the plasma membrane, may provide insights into their mechanisms of action. A topology model allows predictions to be made about which protein regions interact by helping to exclude the possibility of certain interactions between regions on opposite sides of the membrane. It also helps reduce the time required to identify protein regions involved in particular functions. As O-antigen glucosylation is thought to take place in the periplasm (Guan *et al.*, 1999), one or more periplasmic region(s) of considerable size must be present for the attachment of a glucosyl residue to the O-antigen. Therefore, topological information can reduce the time required to identify protein regions involved in particular functions. By identifying such regions, it can provide a basis for their characterisation and localisation of the active site. In addition, topological information enables us to make direct comparisons to other well characterised proteins for structural homology, which may give greater insights into their function.

A dual reporter system consisting of alkaline phosphatase (*phoA*) which is in-frame with the β-galactosidase α-fragment (*lacZα*) developed by Alexeyev and Winkler (1999) was used to help determine the topology of GtrIV. Alkaline phosphatase (AP) is an *E. coli* enzyme that catalyses the hydrolysis of a variety of phosphomonoesters in the periplasm. It is only active when located in the periplasm. This is based on the assumption that in the periplasm, the mature part of PhoA is oxidised such that the cysteine residues are able to form disulfide bridges which enables PhoA to fold correctly (Alexeyev and Winkler, 1999). The process of folding and the assembly of PhoA occurs only after export to the periplasm. In this case, since its signal sequence has been removed,

localisation depends on the protein region to which it is fused (Sugiyama *et al*, 1991; Alexeyev and Winkler, 1999).

In contrast to AP, β-galactosidase (BG) is only active in the cytoplasm. BG is encoded by the *E. coli lac* operon and splits lactose into glucose and galactose. Its activity is only restricted to the cytoplasm as the α-fragment must interact with the cytoplasmically located ω-fragment to cause α-complementation (Alexeyev and Winkler, 1999) Thus fusions to periplasmic sites will be inactive whereas fusions to cytoplasmic domains will be active.

The enzymatic activity of AP can be detected by using 5-bromo-4-chloro-3-indolyl phosphate (Xphos). This substrate is cleaved by AP to produce a blue colouration. On the other hand, the enzymatic activity of BG can be detected by 6-chloro-3-indolyl-β-D-galactoside (Red-Gal) which is cleaved by BG to produce a red colouration. The use of dual indicator (DI) plates containing the substrates of AP and BG discriminates between non-informative (for example out-of-frame) fusions that are colourless (described as white), red cytoplasmic fusions and blue or even purple periplasmic fusions. For both enzymes, there are more quantitative assays based on different absorbance properties of *p*-nitrophenyl phosphate upon hydrolysis by AP and the cleavage of *o*-nitrophenyl-β-D-galactoside by BG that have been described (Manoil *et al.*, 1990). These assays were done simultaneously to quantitatively determine the enzymatic activity that takes place at each fusion as reported by Alexeyev & Winkler (2002) who stated that by using the PhoA/LacZ dual reporters, BG and AP activities can be measured at the same point simultaneously without resorting to genetic recombination to switch fusions. The normalized activity ratios (NAR) can correct for variable expression. Thus, determination of protein synthesis rates is not necessary.

3.2 Previous work on GtrIV

In the past, two approaches were carried out to experimentally determine the topology of GtrIV (Anesh Nair, Honours Thesis 2006). The first approach was by using exonuclease III (Exo III) deletion from the end of the *gtrIV* gene followed by in-frame fusion with *phoA/lacZ.* This method developed by Sugiyama *et al.* (1991), enables a rapid creation of random fusions that span the entire length of *gtrIV* successfully. Another method was also carried out involving a PCR-based approach to create C-terminal deletion fusions at predetermined points.

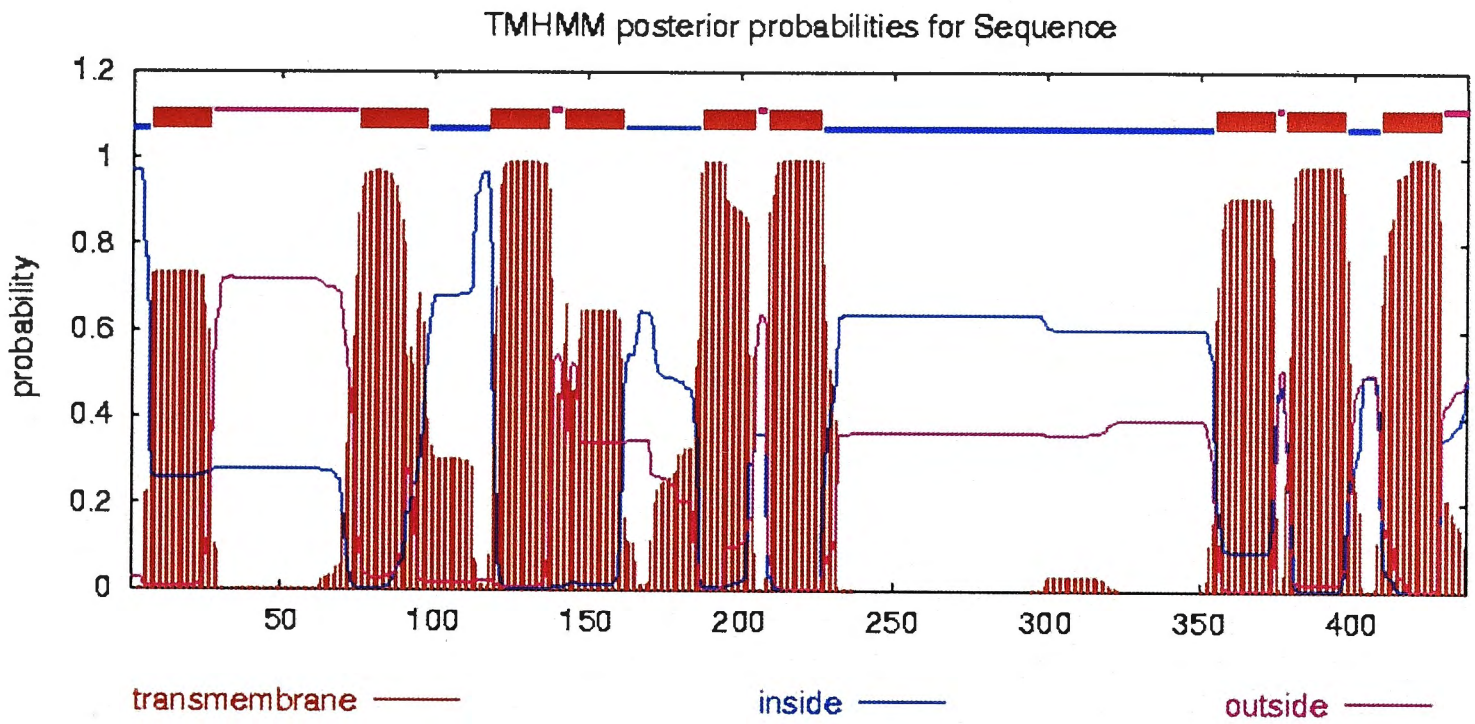
A predicted topology model for GtrIV was first created based on the results given by the various computer prediction programs. The strategies used to identify transmembrane helices and the predictions made by each program are presented in Table 3.1. The DAS and TMHMM predictions are also shown in Figure 3.1. Based on the results of the computer predictions, a consensus model was constructed. In this model, GtrIV was depicted to have ten transmembrane helices, a cytoplasmic N-terminus, a small cytoplasmic C-terminal end, and two large periplasmic loops between helices I and II, and VII and VIII (Figure 3.2).

To confirm this model, Exo III deletion and PCR-based fusions were then carried out to create *gtrIV-phoA/lacZ* truncations within the *gtrIV*. pNV1473 was the template used (Figure 3.3). It was engineered such that *gtrIV* is in tandem with the dual reporter. This construct also contained a restriction site that leaves a 5' overhang susceptible to Exo III digestion closest to the end of *gtrIV* and a restriction site that generates an Exo-resistant 3' overhang just upstream of *phoA/lacZ* to protect the reporter. The random C-terminal fusions brought about by Exo III deletion resulted in 14 blue fusions that were confirmed by the AP and BG assays to be in the periplasm due to their NAR values. All these fusions occur in the periplasmic loops No 2 and No 8. As such, these results

Table 3.1: Results from the computer programs used to predict GtrIV topology and the basis for each prediction.

Program	Basis for prediction	Reference	N-terminus	Number of predicted transmembrane helices
HMMTOP	Hidden Markov Model	Tusnady and Simon (1998)	In	8
SOSUI	Hydropathy profile	Hirkawa *et al.* (1998)	-	10
TMPred	Comparisons with database of known transmembrane proteins	Hofmann and Stoffel (1993)	In	8
DAS	Dense Alignment Surface Method	Czerzo *et al.* (1997)	-	11
TopPred	'Positive inside' rule	von Heijne (1992)	Out	9
TMHMM	Hidden Markov Model	Sonnhammer *et al.* (1998)	-	9

a)



b)

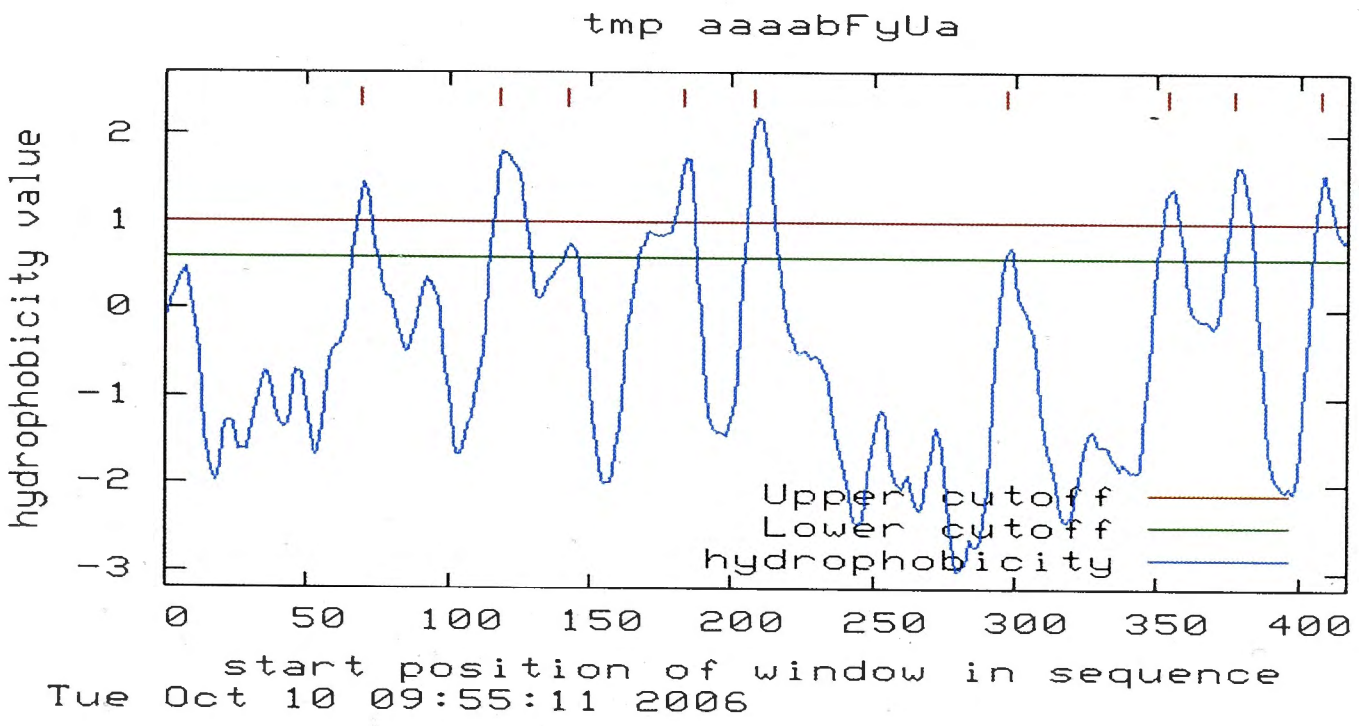


Figure 3.1: Representative computer-based topology predictions of GtrIV. (a) TMHMM (9 helices). The red peaks indicate transmembrane regions, the blue lines indicate cytoplasmic regions and the pink lines indicate periplasmic regions. (b) DAS (9 helices). The peaks correspond to predicted transmembrane helices.

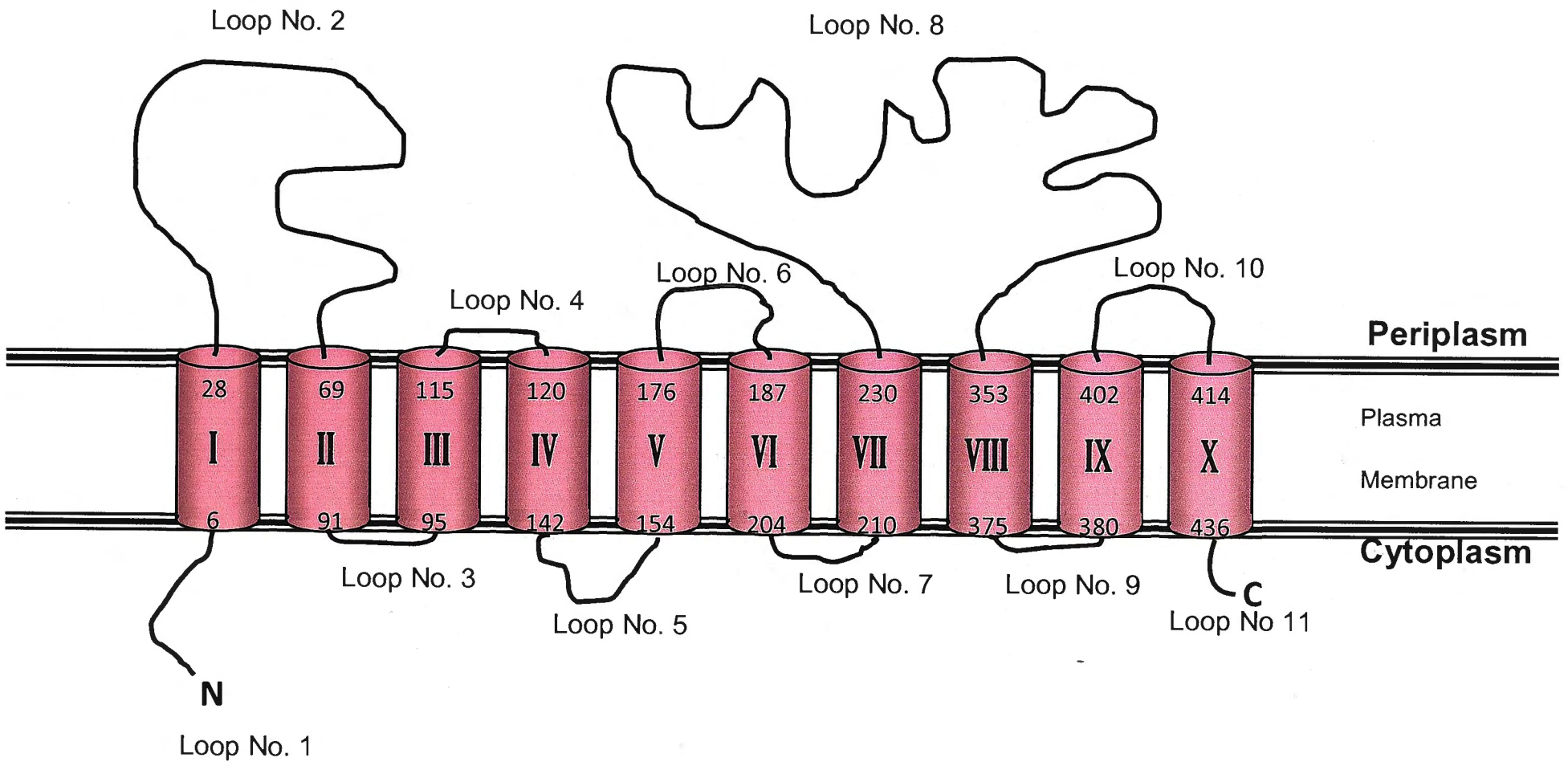


Figure 3.2: The consensus topology of GtrIV. In this model, GtrIV is predicted to have 10 transmembrane helices, two large periplasmic loops and short N-terminal and C-terminal tails

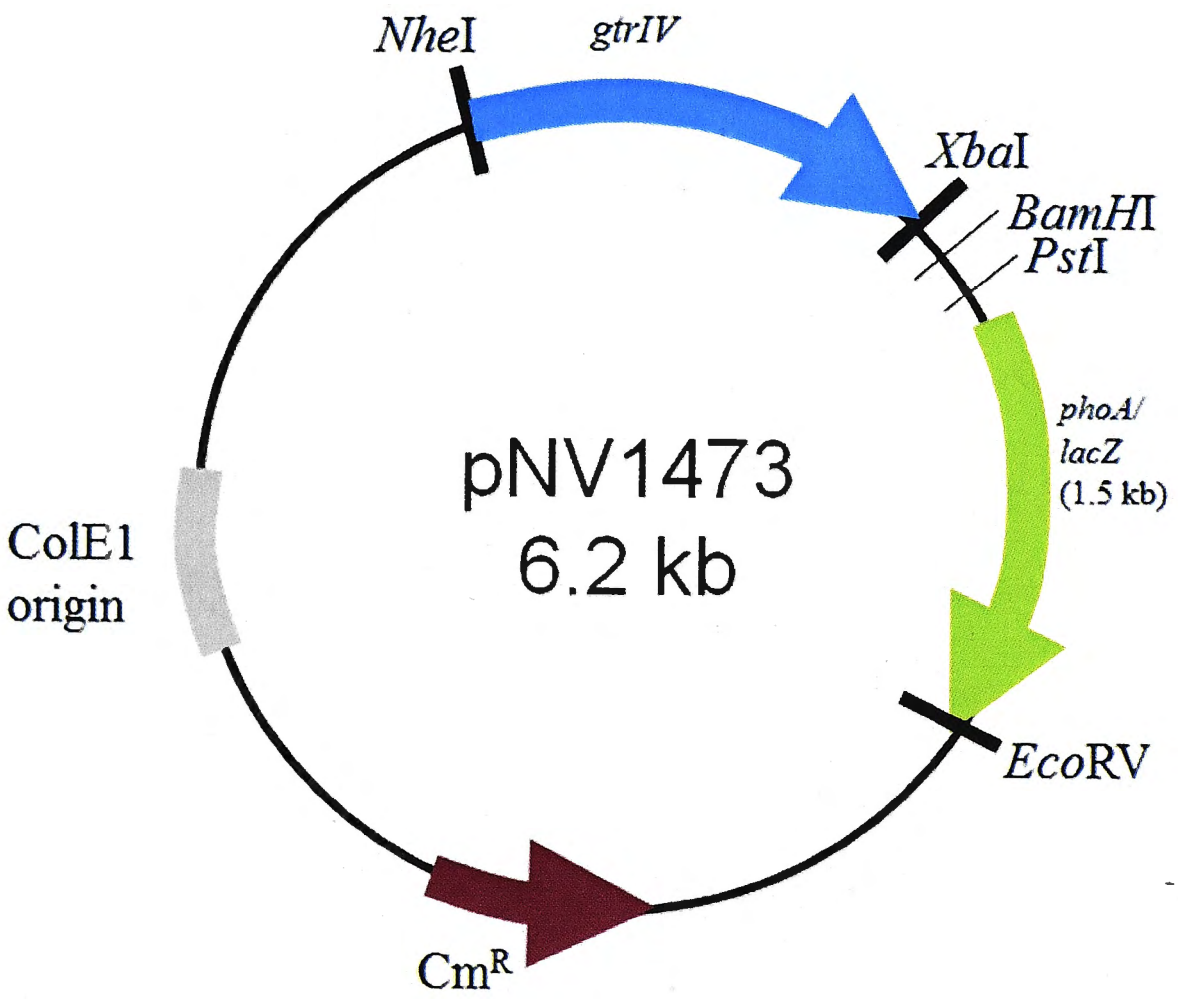
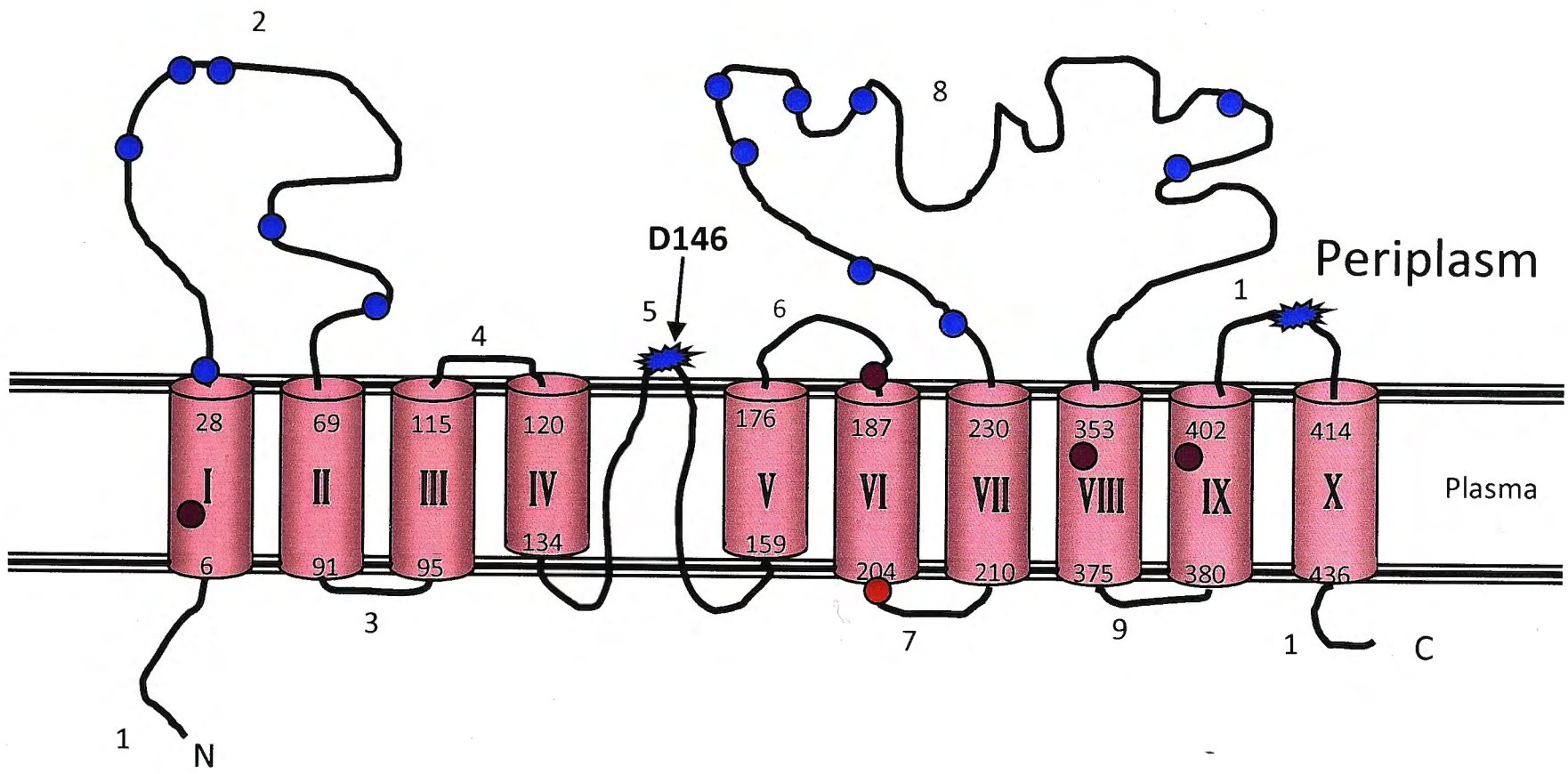


Figure 3.3: The structure of pNV1473 depicting *gtrIV* and *phoA/lacZ* in tandem. The *BamH*I site close to the end of *gtrIV* provides a 5′ overhang susceptible to Exo III while *Pst*I site close to the start of *phoA/lacZ* leaves an Exo-resistant 3′ overhang. Cm^R = Chloramphenicol Resistance

showed that there were no fusions in the two loops that conflicted with the consensus model of GtrIV. Only one in-frame red fusion was obtained by Exo III deletion. A series of PCR-based fusions were then carried out to further confirm the consensus model. However, this approach led to the creation of a blue fusion, D146, which conflicted with the model, as it indicated that D146 is located in the periplasm instead of the cytoplasm. To this accord, a second model was put forward (Figure 3.4). In this model, in order to satisfy all fusions, the presence of a hypothetical re-entrant loop between transmembrane helices IV and V was proposed. As the presence of a re-entrant loop was hypothesised based on just one fusion, it could not be relied on completely and further fusions had to be created to confirm its presence. All fusions that were created in Figure 3.4 were done during my honours year. From this point onwards, all the subsequent fusions were created as part of my PhD project. Once all the fusions were obtained (including the ones from this PhD project), the AP and BG assays (described later in this chapter) were carried out on them. Thus the AP and BG assays were repeated on the fusions created during honours such that all the fusions could be taken as one dataset. The results of the enzyme assays carried out in honours are not used in this thesis.

3.3 Creation of *gtrIV-phoA/lacZ-gtrIV* sandwich fusions

So far, fusions around the region where the proposed re-entrant loop is located have been elusive as none were obtained using Exo III deletion. To circumvent this, two approaches were adopted. These two methods were the use of sandwich fusions and the PCR-based fusion technique. The construction of sandwich fusions within GtrIV can indicate a more accurate representation of topology since the whole protein is present with the dual reporter sandwiched in the middle. A basic overview of how sandwich fusions are created is illustrated in Figure 3.5. Four protein loops that were not covered using the Exo deletion approach were targeted for sandwich fusions: those



Blue periplasmic ExoIII fusions that exhibit Alkaline phosphatase activity

Red cytoplasmic ExoIII fusions that exhibit β-galactosidase activity

Purple ExoIII fusions occurring either in the transmembrane helices or adjacent to them. These fusions exhibit activity for both Alkaline phosphatase and β-galactosidase.

Blue PCR-constructed fusions that exhibit Alkaline phosphatase activity

Red PCR-constructed fusions that exhibit β-galactosidase activity

Figure 3.4: The adjusted topology model of GtrIV which include the ExoIII fusions and the PCR-constructed fusions. Due to the blue fusion at D146, the model was adjusted to include a hypothetical re-entrant loop between transmembrane helices IV and V. This re-entrant loop is only hypothetical until in-frame fusions can be obtained for the cytoplasmic and periplasmic loops on either side of the putative re-entrant loop.

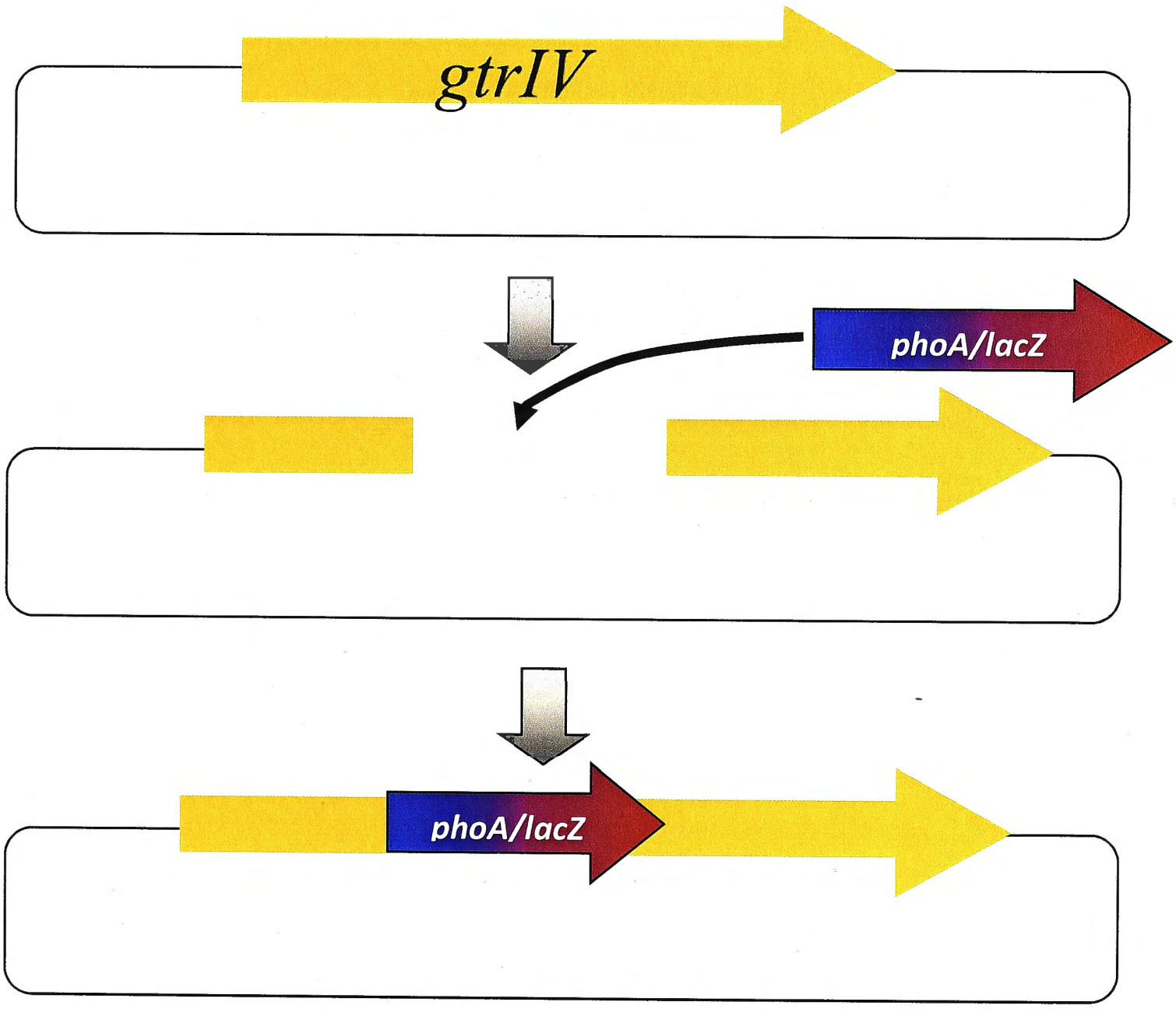
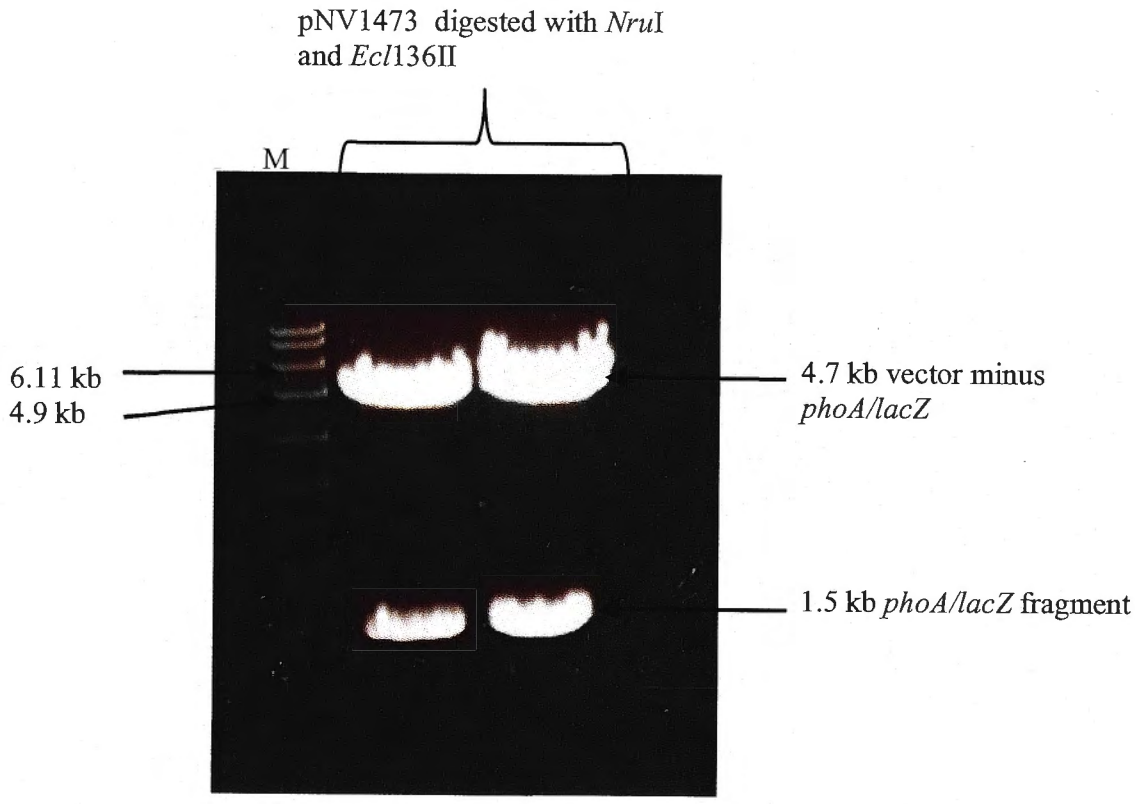


Figure 3.5: Strategy used for sandwich fusions. The vectors used to create these fusions were derivatives of pNV1914 (pBC SK containing *gtrIV* with *phoA/lacZ* excised out). The plasmids were linearised via *Nru*I digests following the introduction of *Nru*I sites within *gtrIV* via site-directed mutagenesis. *phoA/lacZ* double-digested from pMA632 is then blunt-end ligated together with the *Nru*I digested vectors. The ligated construct now contains *phoA/lacZ* sandwiched in-frame with *gtrIV*.

between putative helices II and III (loop No. 3), III and IV (loop No. 4), V and VI (loop No. 6), and VIII and IX (loop No. 9).

Before the sandwich fusions could be made, however, a template had to be constructed that only contained *gtrIV* and not *phoA/lacZ*. As such, *phoA/lacZ* was excised from pNV1473 by carrying out a double digest with *Nru*I and *Ecl*136II (Figure 3.6). The resulting 4.7 kb vector was cut out of the agarose gel, column purified and religated. The ligation mix was transformed into electrocompetent XL1-blue cells and plated out onto agar plates containing chloramphenicol. The resulting clones were screened via *Xba*I restriction digests. Following this, one out of 7 positive clones was chosen and sequenced for confirmation. This new construct, pNV1914, would now be used to create the sandwich fusions. The *Nru*I sites were then introduced into the *gtrIV* sequence at the points of interest using site-directed mutagenesis (pNV1914 as template). The primers used were GtrIVR93NruIF and GtrIVR93NruIR (to target the loop between helices 2 and 3 at residue R93), GtrIVK117NruIF and GtrIVK117NruIR (loop between helices 3 and 4 at residue K117), GtrIVA181NruIF and GtrIVA181NruIR (loop between helices 5 and 6 at residue A181), and GtrIVR375NruIF and GtrIVR375NruIR (loop between helices 8 and 9 at residue R375). The changes made to the *gtrIV* DNA and amino acid sequences were as conservative as possible. The resulting PCR products were run on gel to check for successful amplification (Figure 3.7a). The mutated constructs were identified using *Nru*I digests and sequenced to confirm the mutation. Retention of *gtrIV* function was confirmed by transforming these constructs into SFL1616 and performing agglutination tests with Type IV antisera (Denka Seiken) to test for serotype conversion from serotype Y to serotype 4a.

a)



b)

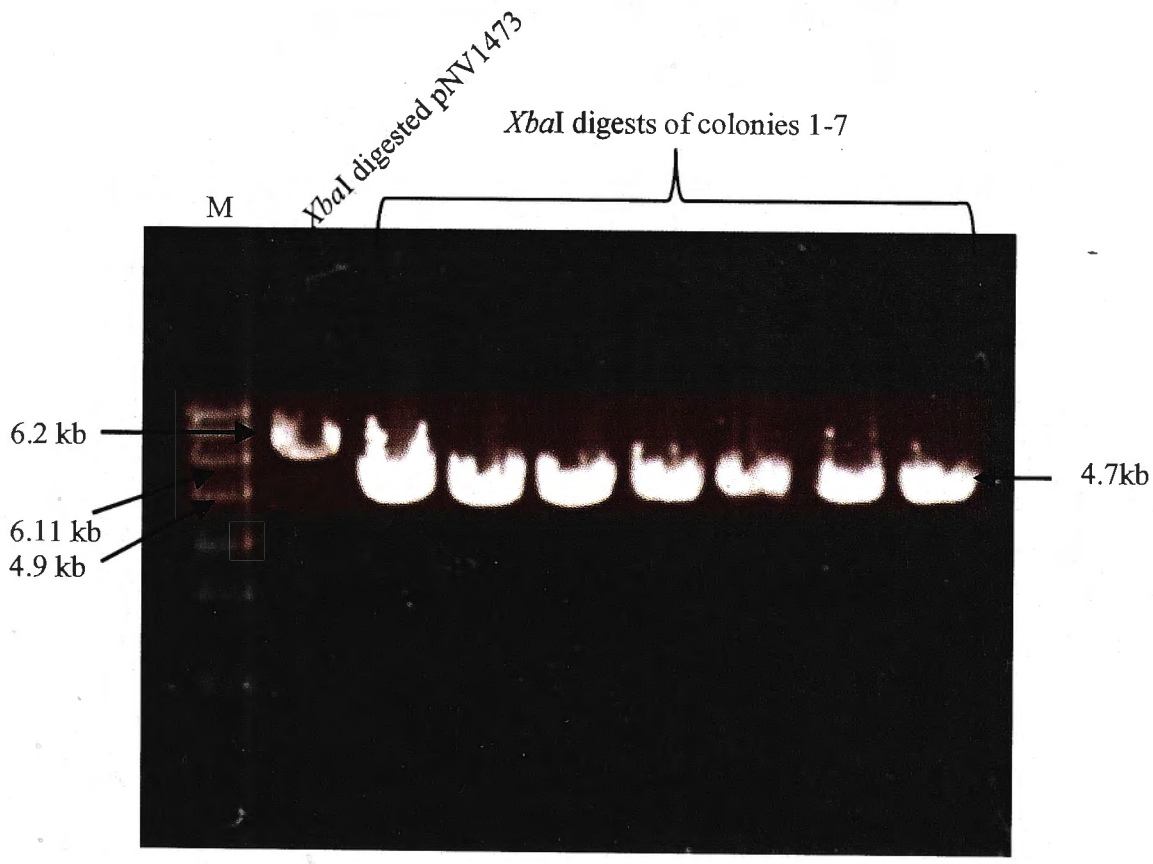
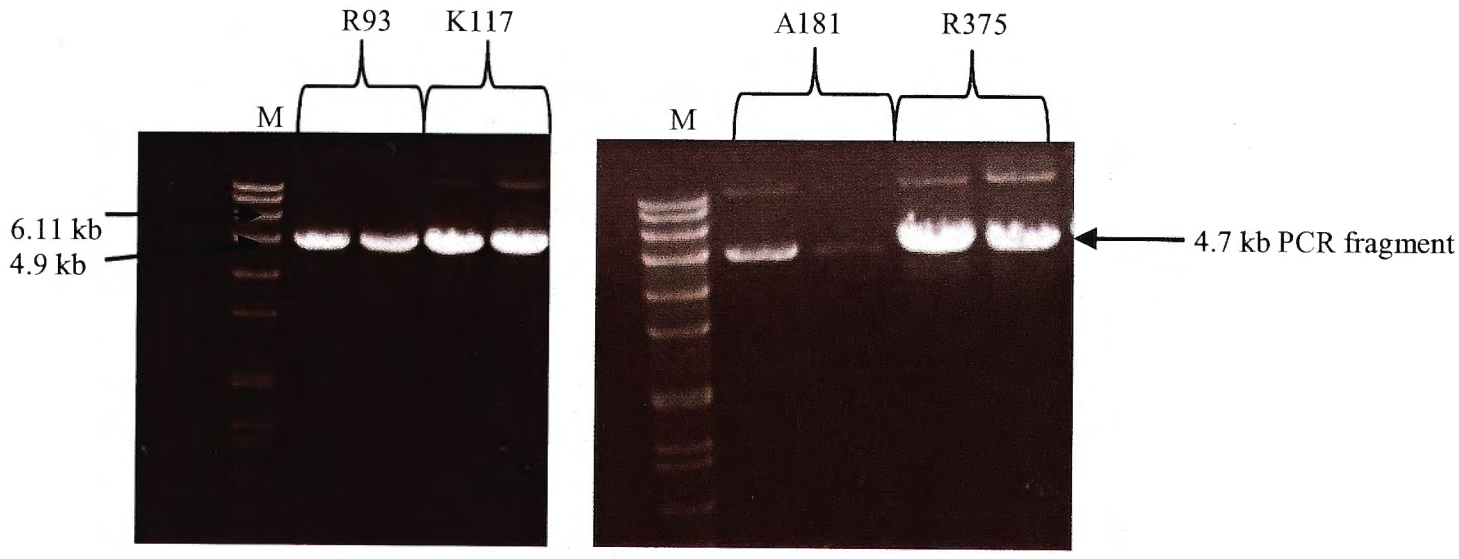
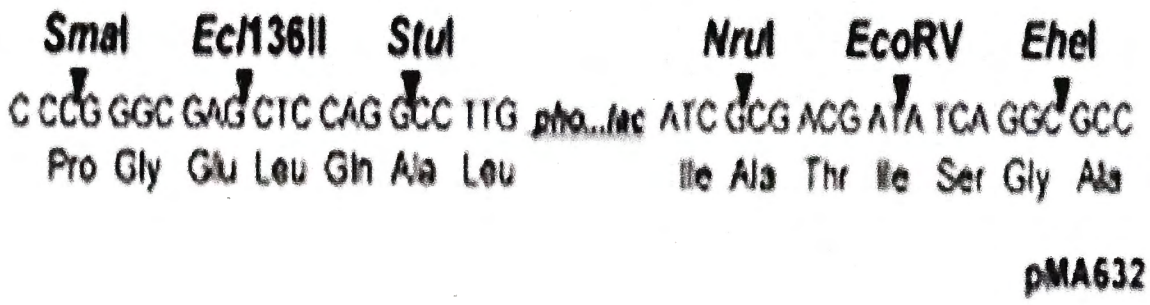


Figure 3.6: Agarose gel electrophoresis depicting the digestion and subsequent screening of pNV1914. a) pNV1743 (*gtrIV* with *phoA/lacZ* downstream and out of frame) was double digested with *Nru*I and *Ecl136*II to leave a 4.7 kb vector fragment and a 1.5 kb *phoA/lacZ* fragment. b) Clones 1 -7 were screened with *Xba*I digests. All 7 clones exhibited a 4.7 kb band. pNV1473 was used as a control and it ran at 6.2 kb. M= Molecular weight marker

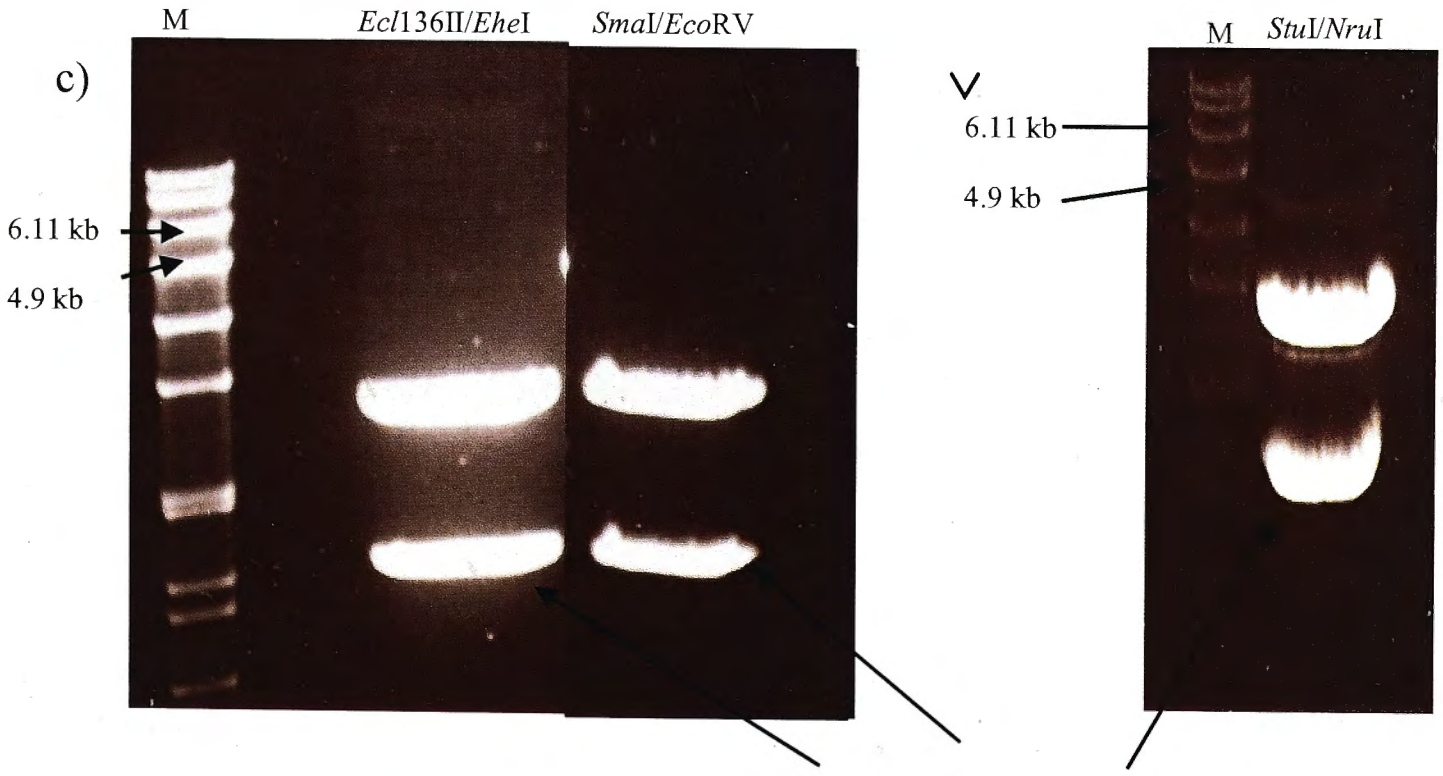
a)

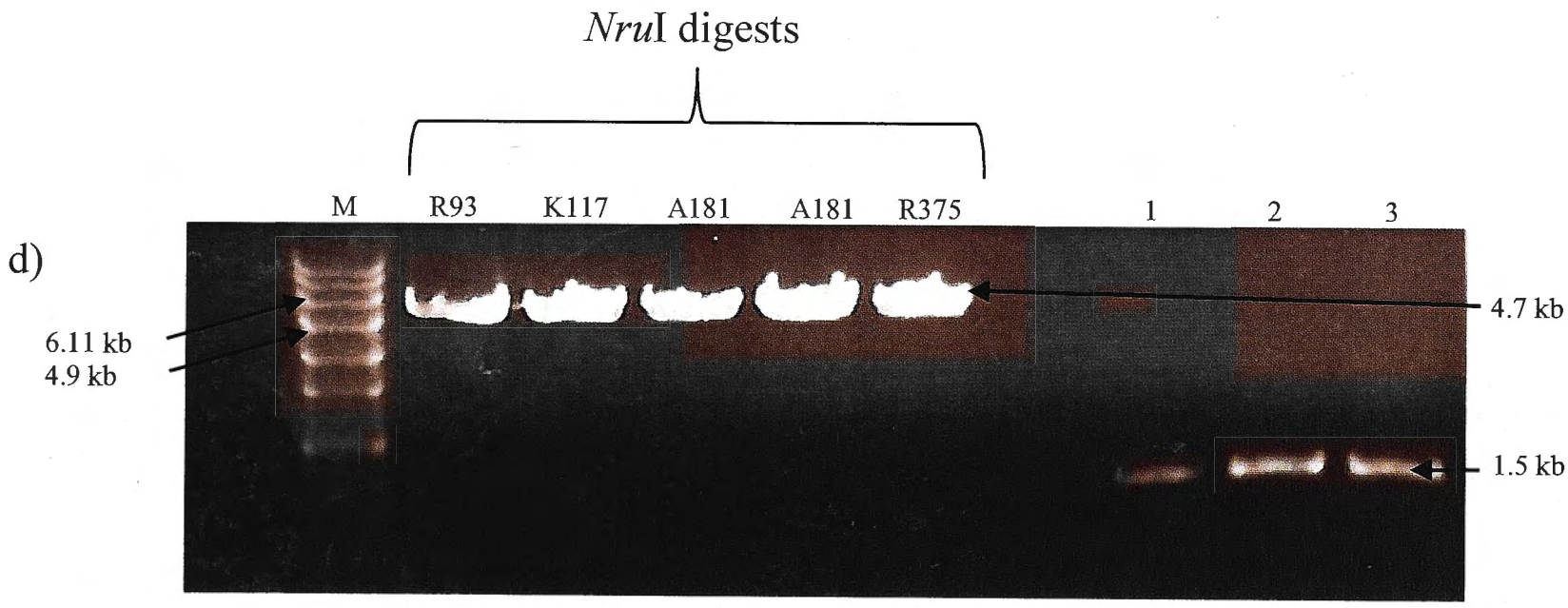

R93
K117
M
6.11 kb
4.9 kb
A181
R375
M
4.7 kb PCR fragment

b)


SmaI
Ecl136II
StuI
NruI
EcoRV
EheI
C CCG GGC GAG CTC CAG GCC TTG pho...lac ATC GCG ACG ATA TCA GGC GCC
Pro Gly Glu Leu Gln Ala Leu
Ile Ala Thr Ile Ser Gly Ala
pMA632

c)


M
Ecl136II/EheI
SmaI/EcoRV
6.11 kb
4.9 kb
M
StuI/NruI
6.11 kb
4.9 kb
1.5 kb phoA/lacZ fragment



1- phoA/lacZ double digested with *Ecl*136II/*Ehe*I
2- phoA/lacZ double digested with *Sma*I/*Eco*RV
3- phoA/lacZ double digested with *Stu*I/*Nru*I

Figure 3.7: a) PCR amplification of sandwich fusion vectors via site directed mutagenesis to insert *Nru*I sites. The samples were labelled according to the corresponding residue where the sites have been inserted. All reactions displayed a 4.7 kb amplicon. b) Sequences flanking *phoA/lacZ* within pMA632. *Sma*I, *Ecl*136II and *Stu*I sites are upstream of *phoA/lacZ* while *Nru*I, *Eco*RV and *Ehe*I sites are downstream of *phoA/lacZ*. c) and d) Double digests of pMA632 to excise *phoA/lacZ* . A combination of *Ecl*136II/*Ehe*I, *Sma*I/*Eco*RV and *Stu*I/*Nru*I double digests yielded 1.5 kb *phoA/lacZ* fragments to be used in for ligation with the amplified vectors containing the *Nru*I sites to create the respective sandwich fusions. d) Purified vectors digested with *Nru*I and the purified double digested *phoA/lacZ* fragments. M= Molecular weight marker.

Upon creation of the mutated vectors, *phoA/lacZ* was then excised from pMA632. In order to maintain the correct frame when inserted in *gtrIV*, 3 different sets of double digests (*Nru*I/*Stu*I, *Ehe*I/*Ecl*136II, and *Sma*I/*Eco*RV) needed to be carried out (Figure 3.7 b and c). All the vectors were then *Nru*I digested while inserts were cut from the gel and purified (Figure 3.7d). The *Nru*I/*Stu*I digested *phoA/lacZ* was blunt-end ligated with the *Nru*I-digested A181 construct. Likewise, the *Ehe*I/*Ecl*136II digested *phoA/lacZ* was blunt-end ligated with *Nru*I digested R93 construct while the *Sma*I/*Eco*RV digested *phoA/lacZ* was blunt-end ligated with *Nru*I digested constructs K117 and R375. The ligation mixes were transformed into JM109 and colonies showing colouration on DI plates were investigated further. The R93, K117, A181 and R375 transformation plates had 200 colonies (30 red), 100 colonies (30 red), 100 colonies (10 blue) and 50 colonies (5 red) respectively. The resulting red and blue colonies were then screened using restriction digests (*Nhe*I) (Figure 3.8). *Nhe*I was chosen as it was one of the restriction enzymes used in cloning *gtrIV* into pBCSK in the first place. Therefore, it is present in all the constructs, including the positive control pNV1914. They were then sequenced using the PhoSeqNewR and M13R primers to confirm that phoA/LacZ was inserted in-frame.

The final constructs bearing the sandwich fusions were pNV1924 (R93), pNV1925 (K117), pNV1926 (A181) and pNV1927 (R375). Sandwich fusions R93, A181 and R375 reported red, blue and red colouration, respectively, and were in agreement with the model. This was confirmed using enzyme assays (Table 3.2). R93 and R375 had NAR of 1: >100 while a181 had a NAR of >100:1. However, K117 was in conflict with this model as it was predicted to be localised within the periplasm but it reported a red colouration and a NAR of 1:>100. This was indicative of a cytoplasmic localisation.

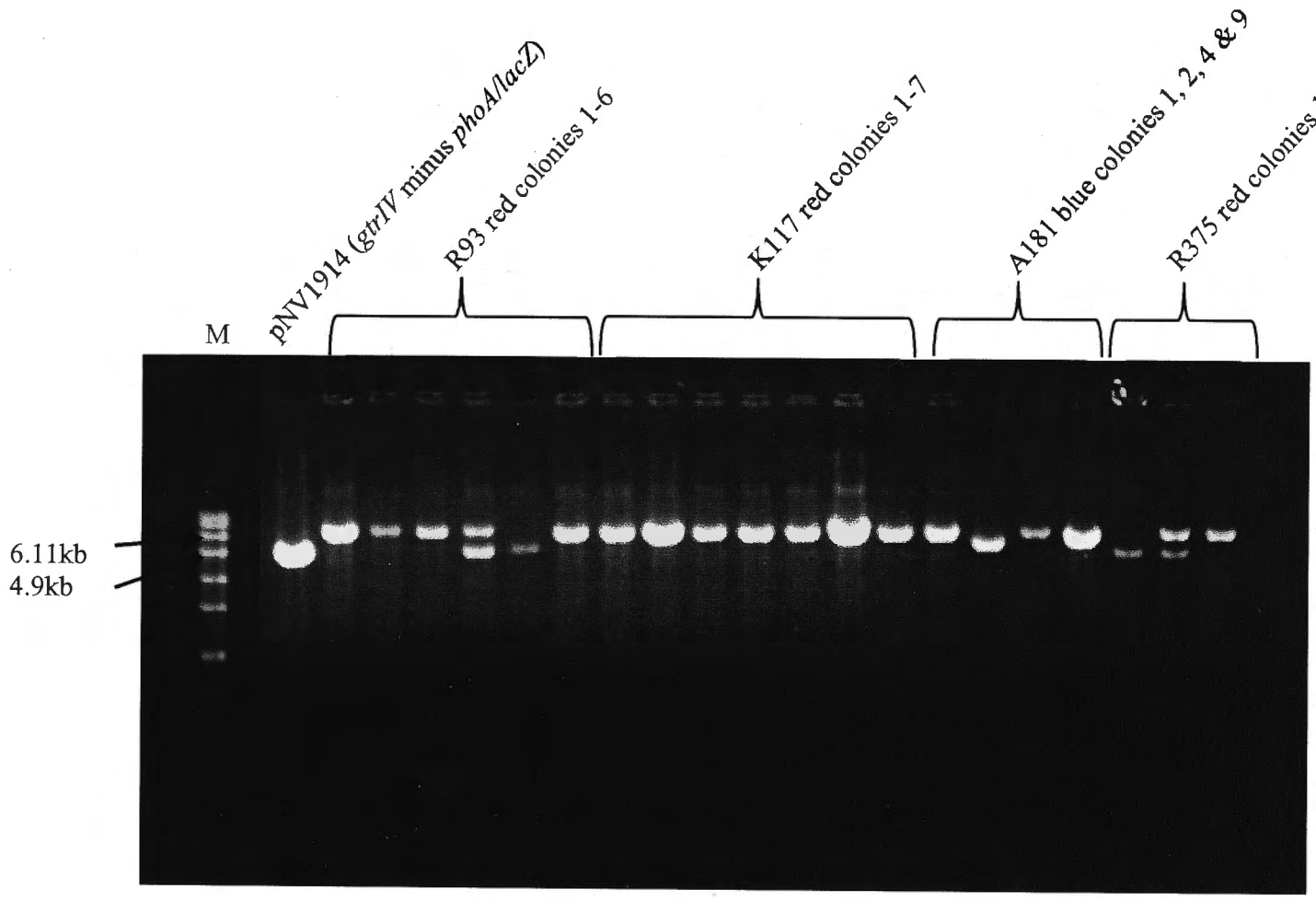


Figure 3.8: *Nhe*I digests of the various red and blue colonies from the sandwich fusion ligations for R93, K117, A181 and R375. Four R93 red colonies displayed a 6.2 kb band. All 7 K117 red colonies had a 6.2 kb band. Three A181 blue colonies, displayed 6.2 kb bands. Only R375 red colony 3 displayed the expected 6.2kb linearized fragment. pNV1914 was also digested and run as a control showing a 4.7 kb band. M= Molecular weight marker

Table 3.2: Analysis of *gtrIV-phoA/lacZ* fusions and *gtrIV-phoA/lacZ-gtrIV* sandwich fusions for GtrIV topology determination.

Sample ID	AA[1]	Colour[2]	Average AP[3]	Average BG[4]	%AP[5]	%BG[6]	NAR[7] (AP:BG)	Location on Model[8]
Random C-terminal Fusions								
B1925	A15	Purple	311 ± 9	25 ± 2	32.6%	16.1%	2:1	t.1
B1910	N29	Blue	356 ± 47	1#	37.3%	0.6%	52:1	p.2
B1924	P38	Blue	357 ± 44	2#	37.5%	1.3%	24:1	p.2
B1907	N41	Blue	296 ± 47	1#	31.1%	0.6%	38:1	p.2
B1935	G42	Blue	195 ± 28	3 ± 1	20.5%	1.3%	20:1	p.2
B1908	L56	Blue	133 ± 29	4 ± 1	14.0%	2.6%	5:1	p.2
B1931	F60	Blue	383 ± 25	1#	40.2%	0.6%	57:1	p.2
B1919	Q186	Purple	9 ± 91	15 ± 2	0.9%	9.7%	1:11	t.4
B1913	I200	Red	22 ± 4	132 ± 6	2.3%	85.1%	1:34	t.4
B1926	N235	Blue	292 ± 51	0#	30.6%	0.0%	>100:1	p.6
B1904	L240	Blue	394 ± 20	1#	41.4%	0.7%	58:1	p.6
B1921	L246	Blue	660 ± 39	1#	69.3%	0.7%	97:1	p.6
B1922	N251	Blue	962* ± 59	1#	100.9%	0.9%	>100:1	p.6
B2338	N255	Blue	553 ± 30	0#	58.1%	0.1%	>100:1	p.6
B1905	D261	Blue	385 ± 55	2#	40.4%	1.2%	33:1	p.6
B1909	S300	Blue	426 ± 21	4 ± 1	43.6%	2.8%	16:1	p.6
B1923	M310	Blue	589 ± 15	1 ± 1	61.8%	0.1%	>100:1	p.6
B1920	V359	Purple	269 ± 34	9 ± 1	28.2%	5.5%	5:1	t.6
B1911	Q393	Purple	12 ± 3	8 ± 2	1.2%	4.9%	1:4	t.7
PCR-Constructed Fusions								
B2340	V102	Red	2#	121 ± 17	0.2%	78.2%	1:>100	c.3
B1937	D146	Blue	953 ± 35	7 ± 1	100.0%	4.3%	23:1	re (p.4)
B2343	V155	Red	6 ± 2	71 ± 9	0.6%	45.8%	1:76	re (c.4)
B2405	K162	Red	5#	11#	0.5%	7.1%	1:14	re (c.4)
B2346	D169	Red	3#	79 ± 17	0.3%	51.1%	1:>100	re (c.4)
B1936	D406	Blue	177 ± 40	1#	18.6%	0.9%	20:1	p.8
B2339	K437	Red	5 ± 1	155** ± 8	0.5%	100.0%	1:>100	c.9
Sandwich Fusions								
B2341	R93	Red	0#	10 ± 2	0.1%	21.3%	1:>100	c.3
B2345	K117	Red	0#	14 ± 2	0.3%	29.4%	1:>100	c.3
B2342	A181	Blue	135+ ± 25	0#	100.0%	0.3%	>100:1	re (p.4)
B2344	R375	Red	1 ± 1	47++ ± 7	0.4%	100.0%	1:>100	c.7

[1] Position of the last residue of GtrIV followed by *phoA/lacZ*, [2] Colour of colony as seen on dual indicator plates, [3,4] Activities of the fusions, average of four independent experiments with standard deviations of each activity, [5,6] Percentages of AP and BG activities measured relative to the maximum activity in the set, [7]Normalised AP:BG activity ratio (NAR) rounded to the nearest integer. [8]Location of the fusion on the adjusted topological model of GtrIV(Figure 1); c, cytoplasm; p, periplasm; t, transmembrane helix; re, re-entrant loop. *Highest truncation AP, **Highest truncation BG, +Highest Sandwich fusion AP, ++Highest sandwich fusion BG. Since the random C-terminal fusions and PCR mediated fusions brought about truncated protein fused to the dual reporter, they were taken as one data set and the highest respective AP and BG values between them was used to calculate the NAR values. The NAR values for the sandwich fusions were calculated separately based on the highest AP and BG values obtained from all the sandwich fusions alone. #Standard deviations were less than 0.5 and not included.

3.4 Creation of PCR-based fusions

In parallel with the sandwich fusions, before a definitive model could be put forth, several PCR based fusions were also created to shed light on the middle portion of GtrIV that contains the putative re-entrant loop. The method of how the PCR-constructed fusions were carried out is shown in Figure 3.9. Using this method, four PCR-based fusions were created (V102, V155, D169 and K43)8. Fusions R93 and K117 indicated that a large cytoplasmic loop may be present between transmembrane helices II and III. As such, a fusion at V102 would confirm this. V155 and D169 were chosen to confirm the presence of a re-entrant loop. A fusion point at K348 was to confirm the cytoplasmic localisation of the c-terminal tail of GtrIV. The reverse primers used to create these fusions were GtrIVV102HpaIR, GtrIVV155HpaIR, GtrIVD169HpaIR and GtrIVCtermHpaIR. In all cases, the template was pNV1493 and the forward primer was PhoFHpaI, which anneals at the beginning of the *phoA/lacZ* sequence. All the primers contained an *Hpa*I restriction site for screening purposes. The resulting PCR products were treated with *Dpn*I to remove template DNA, run 0.7% agarose gel to confirm for successful amplification (Figure 3.10) and column purified. The purified DNA was then digested with *Hpa*I, column purified once more and then self-ligated. The ligation mixes for each fusion were transformed into electrocompetent JM109 cells and plated out onto agar plates containing chloramphenicol. The resulting clones were screened via *Hpa*I digests and sequenced with PhoSeqNewR primer to confirm that the fusions were in-frame. Upon confirmation of the fusions, they were then patched onto DI plates to observe for colour.

The PCR-constructed K437 displayed a red colour along with a NAR of 1:>100 that confirms that the C-terminal end is cytoplasmic. Truncation fusion V102 reported a red colouration on DI plates and a high BG NAR compared to AP (1:>100). This fusion was in between both sandwich fusions R93 and K117 that displayed NAR values and colouration favouring a higher BG

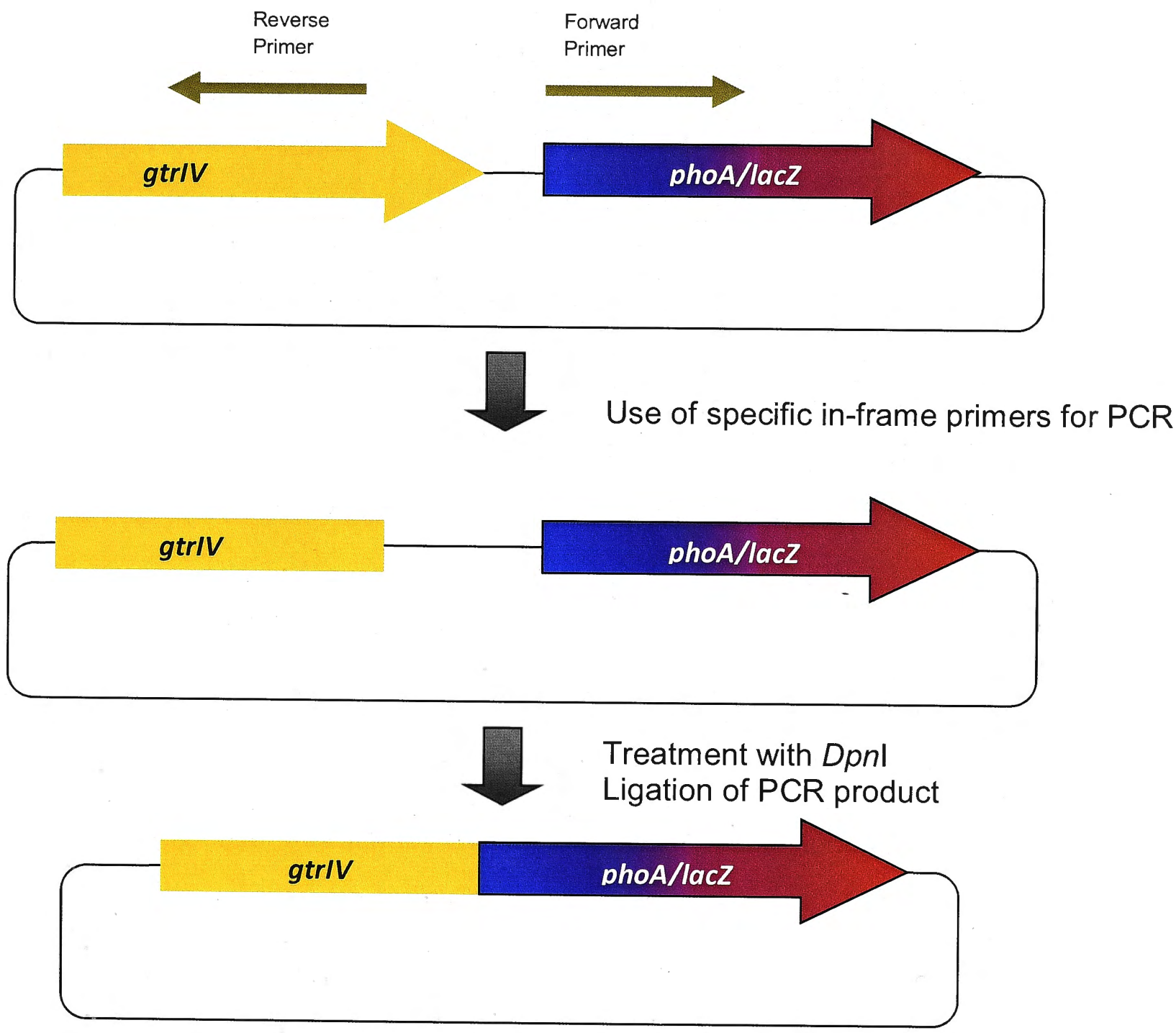


Figure 3.9: Strategy used for PCR-constructed fusions. The template used for this amplification was pNV1473. The forward primer used for PCR was PhoF, which anneals at the beginning of the *phoA/lacZ* sequence. In-frame reverse primers specific for each desired fusion were used.

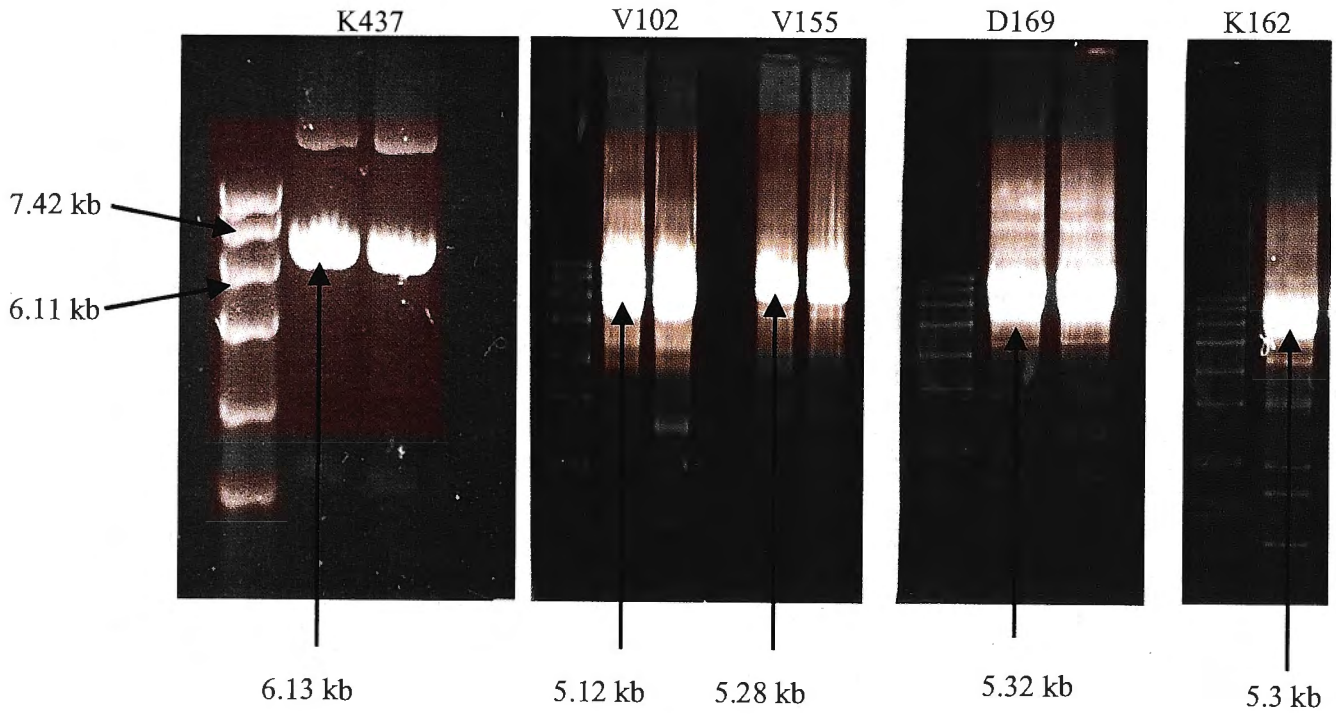


Figure 3.10: PCR-based fusion reactions for K437 (C-terminal fusion), V102 (loop No. 3), V155 (re-entrant loop), D169 (re-entrant loop). The amplified PCR product for K438 runs at ~6.13kb while that of V102 runs at ~5.2kb.V155 and D169 display similar sized bands (5.28 kb and 5.32 kb, respectively). A PCR-based fusion at K162 was later carried out to help clarify the topology of the re-entrant loop region. The PCR product showed a band of 5.3 kb in size. M= Molecular weight marker

ratio to AP and a cytoplasmic localisation. According to these results, the model was modified accordingly such that transmembrane helix III from the consensus mode (Figure 3.2) was pushed into the cytoplasm along with periplasmic loop No. 4. This consequently would invert transmembrane helix IV, thus allowing the consensus loop No. 5 (containing D146) to be localised in the periplasm. However, inverting loop No. 5 into the periplasm also meant that a large periplasmic loop between the modified transmembrane helices III and IV spanning 47 amino acids was created by including transmembrane helix V of the consensus model. Truncation fusions V155 and D169 were evenly spaced between D146 and A181 and helped to verify this. Both fusions displayed red colouration and NAR values 1:76 and 1:>100, respectively. In order to satisfy the current results, a re-entrant loop was introduced in the proposed model that completely traverses the plasma membrane into the cytoplasm and loops back up into the periplasm such that V155 and D169 are both localised in the cytoplasm. The reasoning behind this is discussed in detail in the Section 3.5 (discussion). Fusion K162, which lies in-between V155 and D169, also reported a red colour on the DI plates. However, it had a relatively low NAR compared to its flanking fusions and the possible reasons behind this are also discussed in Section 3.5 (discussion).

Based on all the fusions and the final NAR values, a finalised topology model was proposed (Figure 3.11). This model shows that GtrIV has 8 transmembrane helices instead of the predicted 10. Both the N-terminal and C-terminal ends are localised in the cytoplasm. GtrIV is seen to have two large periplasmic loops namely, loop No. 2 and loop No. 6. These two loops are hypothesised to be catalytically active during glucosylation. The model also shows the presence of a re-entrant loop that starts at the periplasmic face transmembrane helix III and traverses down to the cytoplasm loops towards the plasma membrane before traversing through the cell membrane back to the periplasm. This model satisfies all fusion points and their respective NAR values.

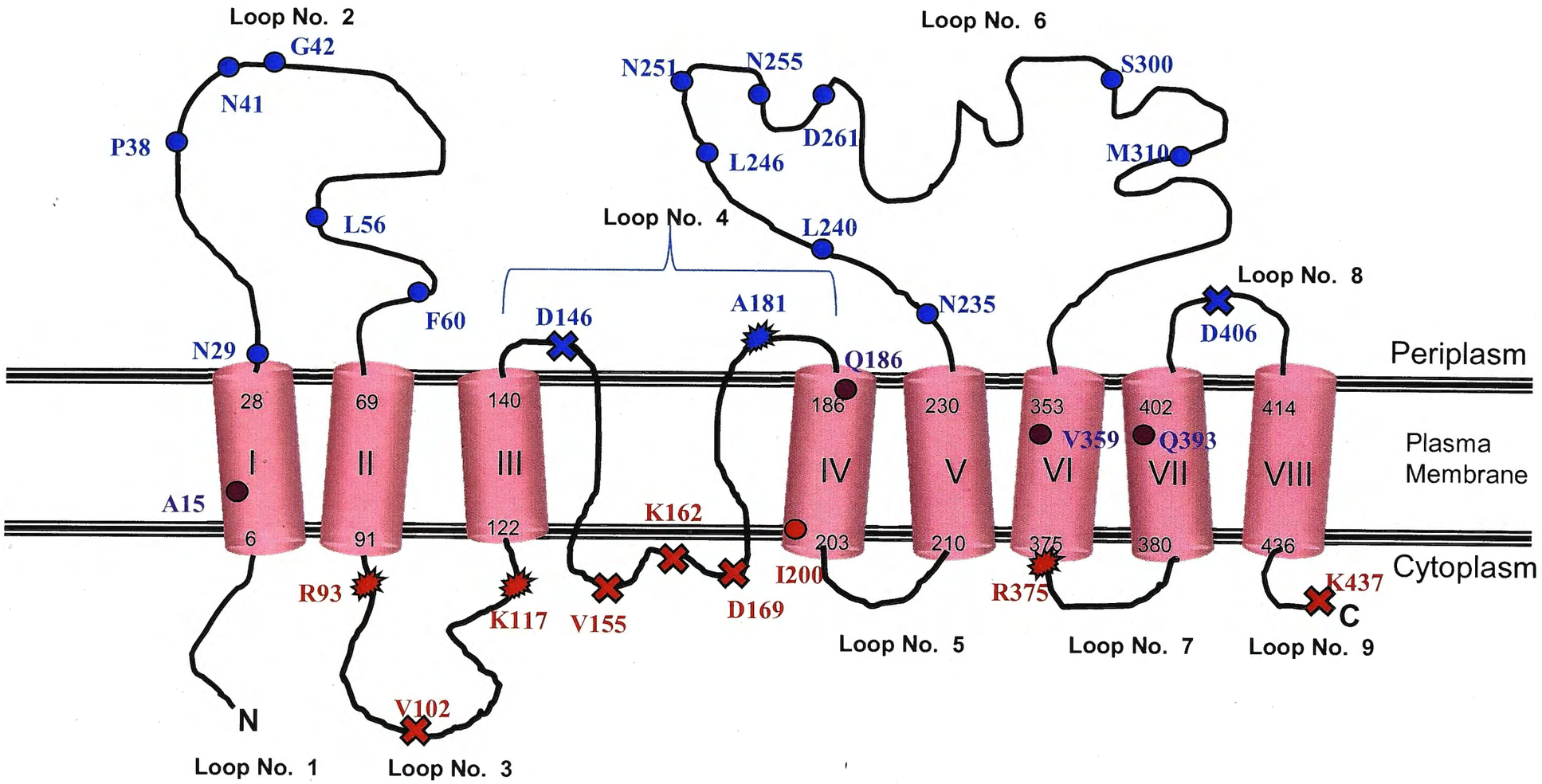

Loop No. 2
G42
N41
P38
L56
F60
N29
Loop No. 6
N251
N255
S300
D261
L246
M310
Loop No. 4
L240
N235
Loop No. 8
D406
D146
A181
Q186
Periplasm
28
69
140
186
230
353
402
414
I
II
III
IV
V
VI
VII
VIII
V359
Q393
Plasma Membrane
A15
K162
6
91
122
203
210
375
380
436
Cytoplasm
R93
K117
V155
D169
I200
R375
K437
C
N
V102
Loop No. 1
Loop No. 3
Loop No. 5
Loop No. 7
Loop No. 9

Figure 3.11: Final GtrIV topology model after the creation of *gtrIV-phoA/lacZ* fusions by the Exo III deletion and PCR-based approaches, and *gtrIV-phoA/lacZ-gtrIV* sandwich fusions. In this model, GtrIV is shown to have 8 transmembrane helices and two large periplasmic loops. All Exo III mediated fusions are indicated with the closed circles while PCR mediated fusions are indicated with the crosses. All sandwich fusions are indicated with the stars. Blue coloured stars, circles and crosses depict fusions that have high PhoA NARs and appear blue on DI plates. Similarly, the ones in red depict fusions that have high LacZ NARs and appear red on DI plates. All purple fusions are seen to have occurred within transmembrane helices. The proposed Re-entrant loop which traverses the plasma membrane from the periplasm into the cytoplasm, forming a cytoplasmic loop, occurs between transmembrane helices III and IV. All fusions have been labelled with the residue at which the fusion with PhoA/LacZ takes place or, in the case of sandwich fusions, where the dual reporter protein has been inserted.

3.5 Discussion

Elucidation of the structure of membrane proteins such as GtrIV and their orientation across the membrane is required in order to make predictions about their mechanism of action. To experimentally verify the hypothetical GtrIV topological model, three different fusion approaches were used. The nested deletion system, used in a previous study for topology determination of GtrIV (Nair A, 2006 Honours Thesis; Nair *et al*, 2011), which utilised exonuclease III (Exo III) deletion from the end of the *gtrIV* gene followed by fusion with *phoA/lacZ,* resulted in about 200 coloured colonies on DI plates. Subsequent screening of these colonies and sequence analysis yielded 19 unique in-frame GtrIV/PhoA-LacZ fusion proteins. The colourations displayed by these 19 fusions were in agreement with the topology model. To further check if these fusions are in agreement with the model, alkaline phosphatase activity (AP) and β-galactosidase (BG) activity for these 19 in-frame fusions were measured via enzyme assays for AP and BG activity and the Normalised Activity Ratios (NARs) were calculated. The enzyme assays were carried out once all sandwich fusions and PCR-mediated fusions had been created and included all fusions (Exo III, PCR-mediated and sandwich fusions) (Table 3.2). The NAR values were generally consistent with the previously observed coloured colonies. Most of the ExoIII mediated fusions were seen to occur in the two large putative periplasmic loops. Given that the consensus model shows that all the cytoplasmic loops are relatively small, there will only be a smaller chance of an in-frame fusion occurring within these loops. According to Alexeyev and Winkler (2002), NARs greater than 2:1 or lower than 1:2 indicate that at least 67% of the reporter activity is properly localised. Alexeyev and Winkler (1999) also stated that the reliability of NARs in providing data about reporter membrane localization is related to the size and diversity of the set of fusions and can be tested by choosing the second highest reference point in calculating the NARs. This was performed (results not shown), and the NARs remained consistent with the previously determined localization of the reporter thus

fortifying our confidence in the finalized topology of GtrIV. All the computer programs did not give the same results and therefore, this highlights the importance of verifying the computer based consensus model. Fusion I200 reported a red DI colouration and a NAR of 1:34. Although this is evidence of cytoplasmic localisation, it showed a high AP activity of 22. Seeing as it also had a substantially high BG activity of 132, it suggests that this fusion is in agreement with the topology model and located in close proximity to the cytoplasmic face of transmembrane helix IV. Alexeyev and Winkler (1999), state that purple fusions result as an association of the fusion to a transmembrane region. This is not an exclusive rule, as seen with fusion I200. The red DI colouration could be a result of the fusion's overwhelmingly higher BG activity coming to the fore. A total of 4 purple fusions were obtained from Exo III deletions. Fusions A15, Q186, V359 and Q393 have NAR values of 2:1, 1:11, 5:1 and 1:4, respectively. Fusion V359 shows periplasmic localisation while fusions Q186 and Q393 show cytoplasmic localisation. Though their NAR values suggest that these fusions are either localised in the periplasm or cytoplasm, it can be seen in table 3.2 that among these four fusions, the lowest average AP activity is 9 while the lowest average BG activity is 8. This therefore indicates that some level of AP and BG activity is present in these fusions. The varying NAR values obtained is attributed to the fact that the %AP values are obtained using highest AP activity of 976, while the %BG values are obtained using the highest BG activity of 155.

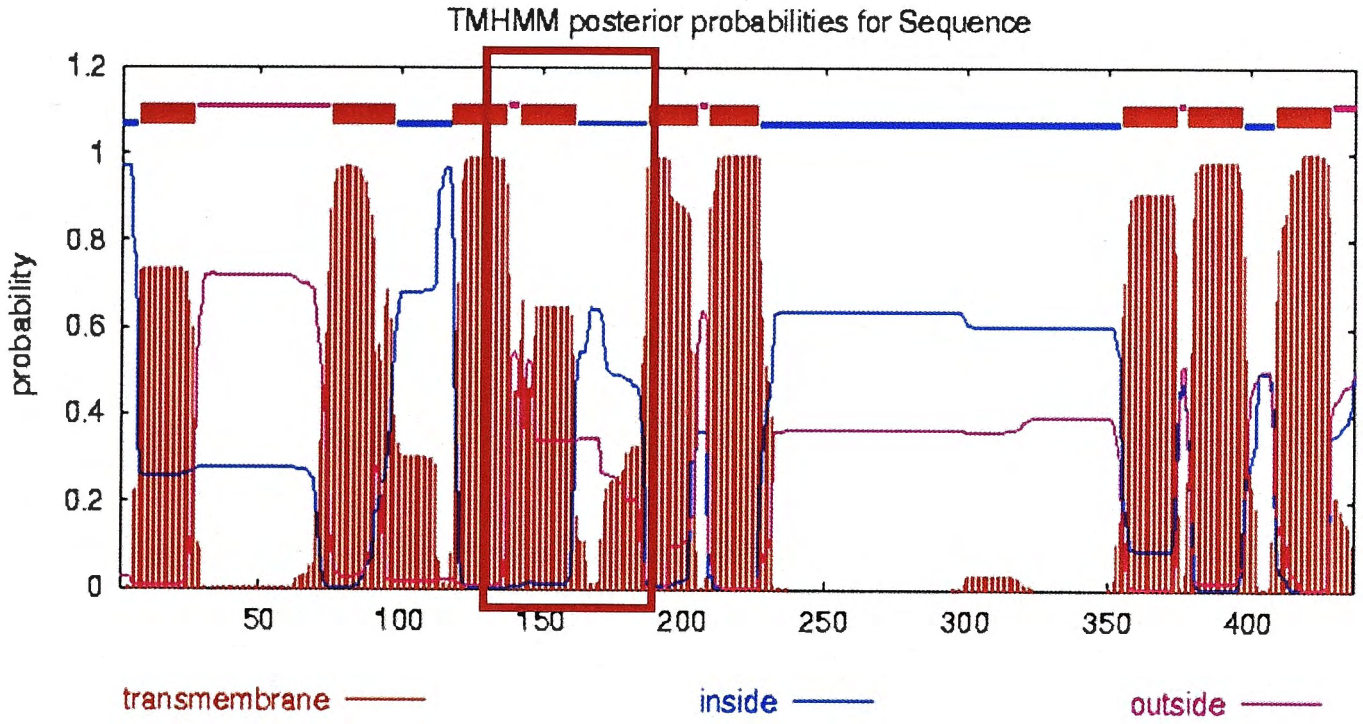
The topology of GtrIV was further elucidated by using two more strategies. These include a PCR-based approach and the use of sandwich fusions. Using the PCR-based fusion technique, it allowed us to exactly fuse the dual reporter to predetermined points in the protein. The second method involved the construction of sandwich fusions. A sandwich fusion can indicate a more accurate representation of topology since the whole protein is present with the dual reporter sandwiched in the middle. These two methods generated a host of constructs. Fusions D146, D406

and K438 were obtained via PCR based fusions. R93, K117, A181 and R375 were generated using *gtrIV/phoA-lacZ/gtrIV* sandwich fusions. A NAR value of 1:>100 for the PCR-constructed red fusion K438 confirms that the C-terminal end is cytoplasmic. Fusion D406 with a NAR value of 20:1 indicated that loop No. 10 (from consensus model shown Figure 3.2) is in the periplasm. While both these fusions satisfied the consensus model, D146 was observed to display blue colouration with a NAR value of 21:1. This fusion was in contradiction to the hypothetical model, which predicts D146 to be localised in the cytoplasm between transmebrane helices IV and V. Sandwich fusion K117 displayed a NAR value of 1:>100 that corresponded with its red colouration and confirms its localisation to the cytoplasm.

TMpred and HMMTOP topology programs predicted similar models in which K117 was predicted to be in the cytoplasm and D146 was predicted to be in the periplasm. Truncation fusion V102 which reported a red colouration on DI plates and a high BG NAR compared to AP, strongly suggests that GtrIV had a cytoplasmic loop between transmembrane helices II and III. This fusion was in between both sandwich fusions R93 and K117 that displayed NAR values and colouration favouring a higher BG ratio to AP and a cytoplasmic localisation. According to these results, the model was modified accordingly such that transmembrane helix III from the consensus mode was pushed into the cytoplasm along with periplasmic loop No. 4. This consequently would invert transmembrane helix IV, thus allowing the consensus loop No. 5 (containing D146) to be localised in the periplasm. However, inverting loop No. 5 into the periplasm also meant that a large periplasmic loop between the modified transmembrane helices III and IV spanning 47 amino acids was created by including transmembrane helix V of the consensus model. To support this theory, truncation fusions V155 and D169 were then designed so that they were evenly spaced between D146 and A181, to verify the newly modelled periplasmic loop. Both fusions displayed red

colouration and NAR values of high BG to AP ratios. In order to satisfy the current results, a re-entrant loop was introduced in the proposed model that completely traverses the plasma membrane into the cytoplasm and loops back up into the periplasm such that V155 and D169 are both localised in the cytoplasm. Analysis of the TMS prediction algorithms indicated that throughout the protein sequence, the transmembrane helices were predicted as clear hydrophobic peaks. However, between amino acid regions P140-I187, two double peaks exist which can be deduced as transmembrane regions (Figure 3.12). This trend was also witnessed for the topology prediction of GtrIc (Ramiscal *et al.*, 2010). The blue fusions obtained via Exo III deletions confirmed the periplasmic locations of loop No. 2 and loop No. 6, while blue truncation fusion D406 confirmed that loop No. 8 (Figure 3.11) is in the periplasm. The presence of two large periplasmic loops gives weight to the hypothesis that glucosylation takes place in the periplasm. Therefore, these two loops provide excellent candidates for GtrIV functional studies. Red fusions I200 and K438 confirm the cytoplasmic localization of loop No. 5 and the C-terminus of GtrI, respectively. Sandwich fusion R375 confirms the cytoplasmic loop No. 7, while sandwich fusions R93 and K117 along with truncation fusion V102 show that loop No. 3 of GtrIV is indeed a large cytoplasmic loop, uncharacteristic to the other Gtrs. Such a large cytoplasmic loop is known to exist in another protein involved in *S. flexneri* O-antigen acetylation, O-acetyltransferase (Oac) (Thanweer *et al.*, 2009). Since glucosylation of the O-antigen is thought to occur in the periplasm, it may play a part in helping to maintain the structural integrity of GtrIV during the glucosylation process. Blue fusions D146 and A181, together with red fusions V155 and D169 suggest the presence of a re-entrant loop between the amino acid regions P140 – I187 that starts from the periplasm and traverses through the plasma membrane into the cytoplasm, forming an intramembrane dipping segment which then traverses back into the periplasm. This unique re-entrant loop structure was put forward following comparisons with GtrIc, the closest structural homologue of GtrIV and is hypothesised to provide structural flexibility in

a)



b)

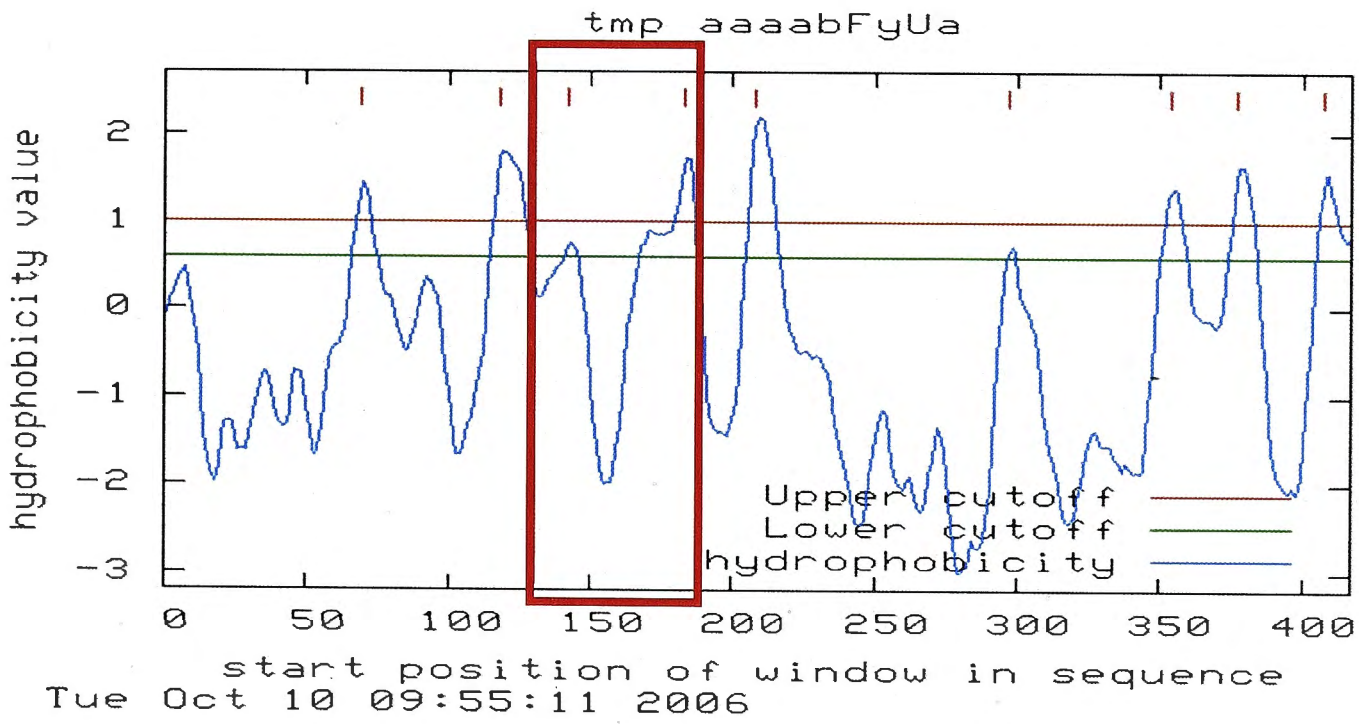


Figure 3.12: Representative computer-based topology predictions of GtrIV. (a) TMHMM. The red peaks indicate transmembrane regions (9 in total), the blue lines indicate cytoplasmic regions and the pink lines indicate periplasmic regions. The red box highlights the doublet peaks that are seen to occur between I136-D187. (b) DAS. The peaks correspond to predicted transmembrane helices (9 in total). As observed with the plot in (a), a doublet peak is observed around the same region and is highlighted with the red box. Such peaks were a common occurrence with the other prediction programs and seem to indicate that a re-entrant loop might be present somewhere within this region.

order to facilitate interaction between periplasmic loop No.2 and periplasmic loop No. 6 (Nair *et al*, 2011). As mentioned before, GtrIc has an almost identical topology to that of GtrIV. Ramiscal *et al* (2010) have also shown that a Gtr-conserved re-entrant loop exists at the lateral centre of GtrIc, between transmembrane helices IV and V. Unlike the re-entrant loops seen in GtrII and GtrV (Korres and Verma, 2004; Lehane *et al*, 2005), Ramiscal *et al* (2010) proposed that it formed a double intramembrane dipping region which can be defined as a protein segment that does not completely penetrate the bilayer but briefly enters and exits the same membrane face (Lasso *et al*, 2006). In light of the current results and considering that GtrIc is the closest homologue to GtrIV, an intramembrane dipping region seems to exist in GtrIV between residues V155 and D169. As this region is rich in isoleucines (hydrophobic residue), it further supports the presence of such an intramembrane dipping region. However, the presence of a KKE tract within this region could also point to a cytoplasmic loop. Therefore, to reinforce this, a final truncation fusion was designed at K162. This fusion reported a red DI colouration and NAR of 1:14. The relatively low NAR value and BG activity of K162 (11) suggests that while in the cytoplasm, this loop might be folded such that it is close to the plasma membrane. There is a possibility of the presence of an intramembrane dipping region occurring within V155 and D169. In this case, the reported red colouration and subsequent BG enzymatic function may have resulted from the effect of PhoA/LacZ weighing down the region such that it is pulled back into the cytoplasm. This could account for its low BG activity as compared to V155 and D169. However, without the use of visual technology such as X-ray crystallography, such structures are difficult to prove. Hence, the presence of a re-entrant loop between transmembrane helices III and IV of GtrIV is proposed. This loop starts at the periplasmic face of the protein and traverses down into the cytoplasm before making its way back to the periplasm. Re-entrant loops have been documented in eukaryotic glutamate transporters (Grunewald *et al*, 1998), in the bacterial potassium channel KcsA (Doyle *et al*, 1998) and in various other

channels as reviewed by Harris-Warrick (2000). This proposed re-entrant loop may be of minimal influence on catalytic potential but may instead provide GtrIV with the conformational flexibility required to transfer a glucosyl group from a membrane lipid to the O-antigen, as proposed by Korres and Verma (2004) for the GtrV protein.

3.6 Conclusion

The present study provides the basis for the functional determination of glucosyltransferase IV involved in O-antigen modification. The topology of GtrIV was experimentally determined by creating different fusions between GtrIV and a dual-reporter protein, PhoA/LacZ. This study shows that GtrIV consists of 8 transmembrane helices, 2 large periplasmic loops, 2 small cytoplasmic N- and C- terminal ends and a re-entrant loop that occurs between transmembrane helices III and IV. Although this topology differs from that of GtrI, Gtr II, Gtr V and Gtr X, it is very similar to that of GtrIc. The discovery of a unique re-entrant loop and the existence of a conserved periplasmic Loop No 2 could help provide an insight in their mode of action in binding and attaching the glucosyl residue to the correct rhamnose of the O-antigen.

Chapter 4

Identifying loops and regions critical for GtrIV function

4.1 Introduction

By identifying critical regions of a protein, it helps provide an insight and may also uncover vital clues to help elucidate its mechanism of action. According to the model of O-antigen glucosylation, the $Gtr_{(type)}$ proteins have been hypothesised to catalyse the transfer a glucosyl group from undecaprenyl phosphate (UndP)-β-glucose to one of the specific sugars of the O-antigen repeat unit (Guan *et al.*, 1999). The subsequent glycosidic linkage formed between the glucosyl group and the sugar is deteremined via interactions of the Gtrs with the acceptor substrates (*N*-acetylglucosamine, RhamnoseI, II and III). Guan *et al.*, (1999) have also postulated that the Gtrs may also be responsible in recycling the free UndP back to the cytoplasmic leaflet of the plasma membrane. It has been hypothesised that each Gtr has a conserved domain and a specific domain. Conserved domains may be involved in interactions with the donor substrate and are responsible for recycling of the lipid carrier. On the other hand, specific domains determine which O-antigen sugar is modified and by what linkage. Two potential segments within all $Gtr_{(type)}$ proteins might be involved in this process of O-antigen modification. These hypothetical segments include: the N-terminal segment responsible for the conserved interaction with the donor substrate UndP-Glucose, and the C-terminal segment responsible for the specific attachment of the glucosyl unit to the specific rhamnose of the growing O-antigen chain.

The newly elucidated topology of GtrIV shows that it contains a large N-terminal periplasmic loop (loop No. 2) and an even larger C-terminal periplasmic loop (loop No. 6) (Nair *et al*, 2011). As O-antigen glucosylation is thought to occur in the periplasm, these two loops are

prime candidates for deletion experiments. As such, series of loop deletion experiments were carried out to investigate their importance to GtrIV function and to help pave the way in identifying critical residues that are situated within the deleted portions of the protein.

Apart from creating deletions within a protein, identification of domains and essential regions for catalysis of membrane proteins can also be achieved by creation of chimeric proteins. This method has been successfully employed in the past to help identify functional domains of various membrane proteins (Buck *et al*, 1994; Giros *et al*, 1994; Nishizawa *et al*, 1995; Cosgriff *et al*, 2000). More recently, Korres and Verma (2006) successfully created chimeric proteins between GtrV and GtrX by swapping their periplasmic loops No. 2. The GtrV-GtrXLoop No. 2-GtrV hybrid retained its native function of converting serotype Y to serotype 5a, thus demonstrating that loop No. 2 of $Gtr_{(type)}$ proteins may have a conservative role in O-antigen glucosylation. As GtrIV and GtrIc have similar topologies, it is postulated that creating chimeras between these two structural homologues would shed some light into the theory that their periplasmic loops play similar roles in their respective functions.

4.2 Identifying regions critical for GtrIV function via loop deletions

4.2.1 Creation of loop No. 2 and loop No. 6 deletion constructs

Multiple alignments of both the large periplasmic loops of GtrIV were carried out against the other Gtrs (see Figures 5.1 and 5.2 in Chapter 5). Although it was found that there was a low level of homology between the periplasmic loops, with the exception of some acidic residues in periplasmic loop No. 2, GtrIV loop No. 2 shares structural homology with the loop No. 2 of

other Gtrs (Korre and Verma, 2004; Lehane *et al*, 2005; Korres and Verma, 2006; Ramiscal *et al*, 2010) (Figure 4.1). Therefore, it is proposed that periplasmic loop No. 2 in all Gtrs may perform the conserved function of interacting with the UndP-Glucose and possibly recycling the lipid carrier back to the cytoplasmic face of the membrane.

To prove the importance of GtrIV loop No. 2 in such function, it was deleted via PCR (Figure 4.2). The template DNA used for this deletion PCR was pNV1917 (encodes for GtrIV that has its C-terminal fused to PhoA/LacZ dual reporter protein) obtained during topology studies. This construct was chosen since GtrIV can be detected on a Western blot using anti-alkaline phosphatase antibodies to determine its expression and assembly in the membrane. The strategy for creating this loop deletion construct is outlined in Figure 4.3. The primers used were IVLp2DelSmaIF and IVLp2DelSmaIR. Both primers contain *SmaI* sites for screening purposes. The resulting PCR products were treated with *Dpn*I to remove template DNA and verified for yield through agarose gel electrophoresis. The *Dpn*I treated PCR products were then purified and digested with *Sma*I. Following this, they were self-ligated and transformed into electro-competent XL1-Blue cells. DNA was extracted from the resulting clones and screened via *Sma*I digests (Figure 4.4a). Clones that were digested by *Sma*I were sent for sequencing to confirm for loop No. 2 deletion. The primers used for sequencing were PhoSeqNewR and A181HpaIR. The plasmid carrying the loop No. 2 deletion was named pNV1918.

Loop No. 6 of GtrIV was found to contain 8 aspartic acid (D) residues and 7 glutamic acid residues (E) (15 negatively charged acidic residues in total) spread throughout the loop, and is hypothesised to be responsible for adding a glucosyl residue to the *N*-acetylglucosamine of the

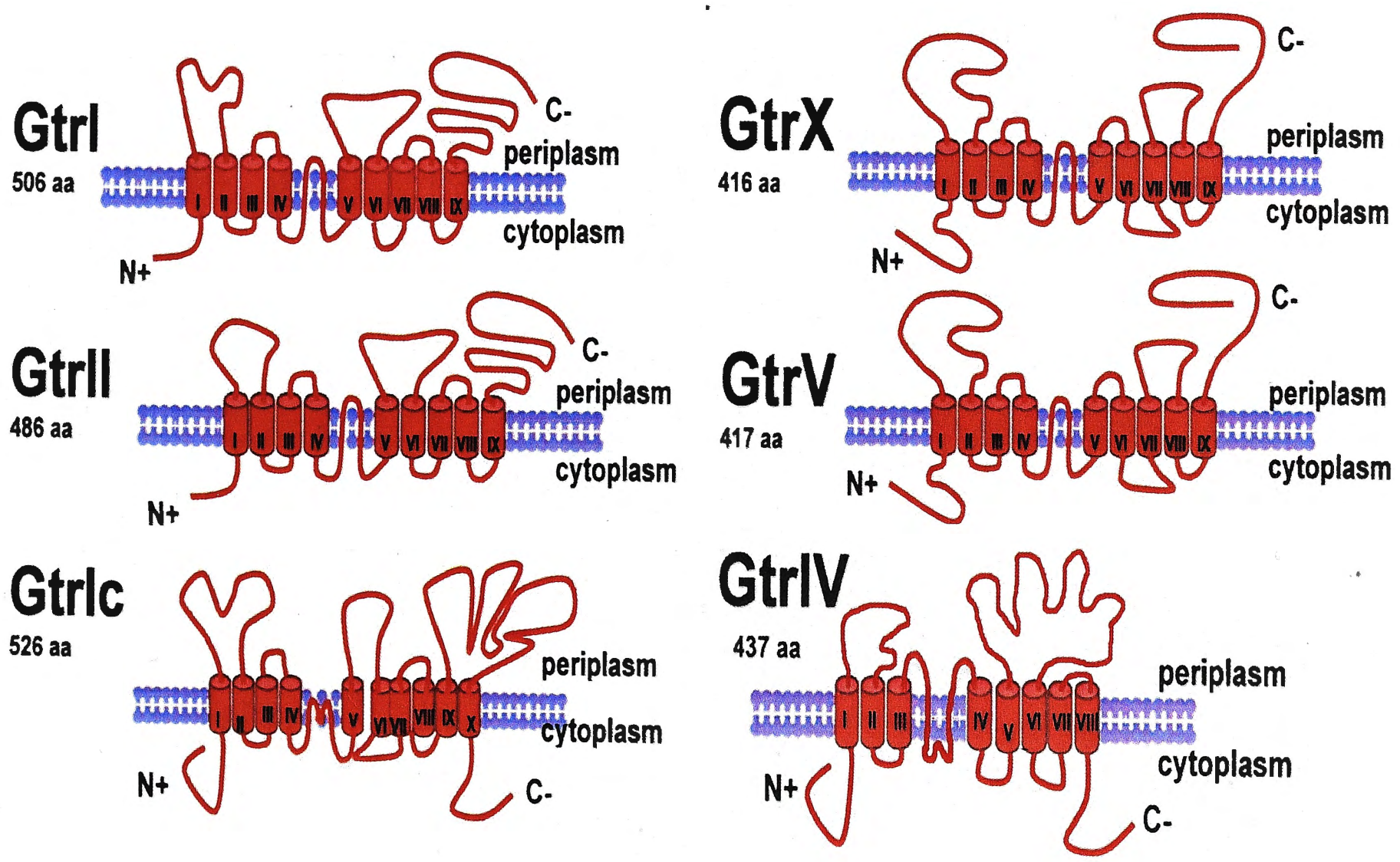


Figure 4.1: Topological comparisons of all the Gtrs. From this diagram, it can be seen that all the $Gtr_{(type)}$ proteins have topologically similar N-terminal periplasmic loops (loop No. 2). However, only GtrIV and GtrIc have large c-terminal periplasmic loops (loop No. 6 and loop no. 10, respectively) and short cytoplasmic C-terminal tails. The rest of the $Gtr_{(type)}$ proteins have long periplasmic C-terminal tails. Adapted from Tran, V (2010) Honours thesis.

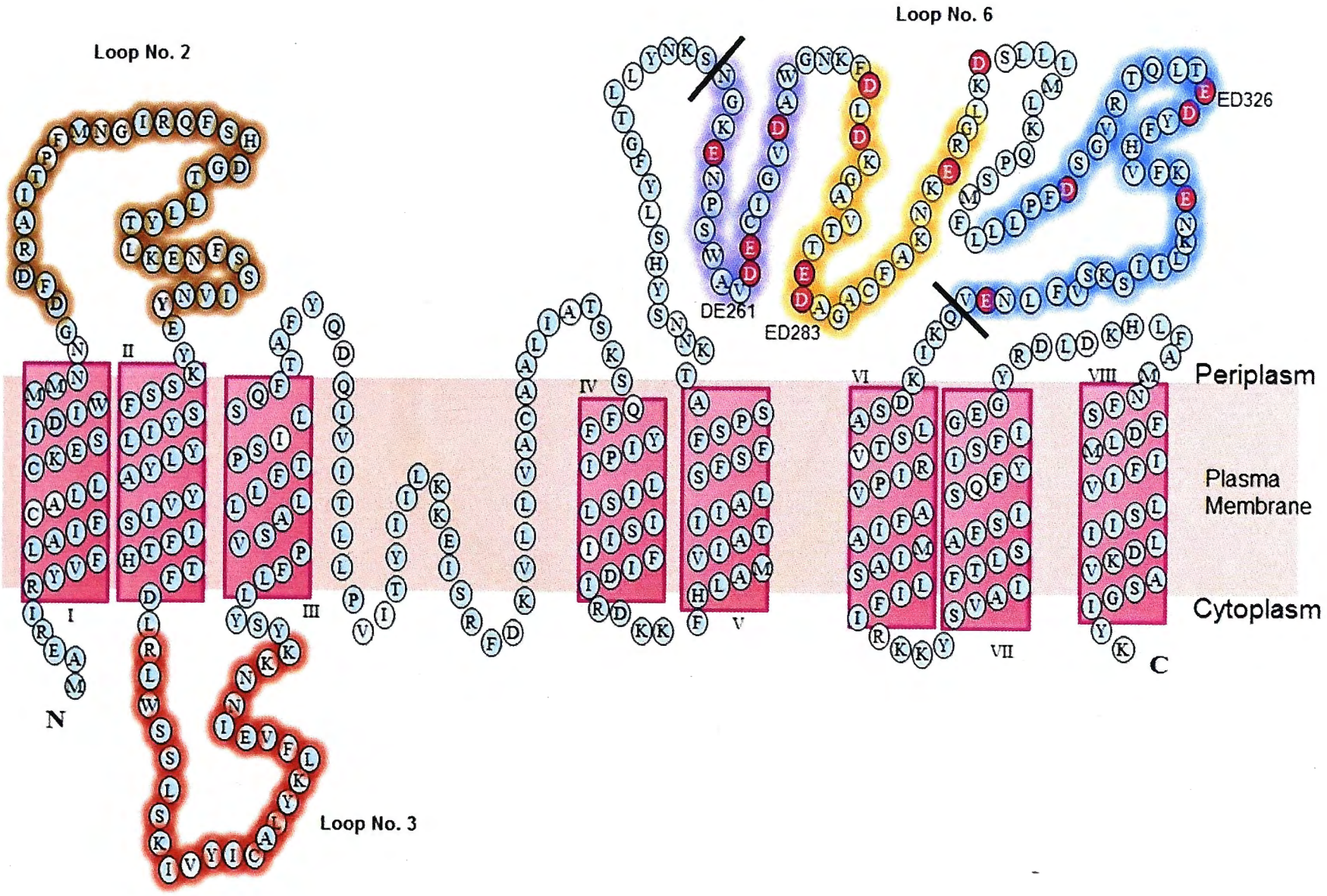
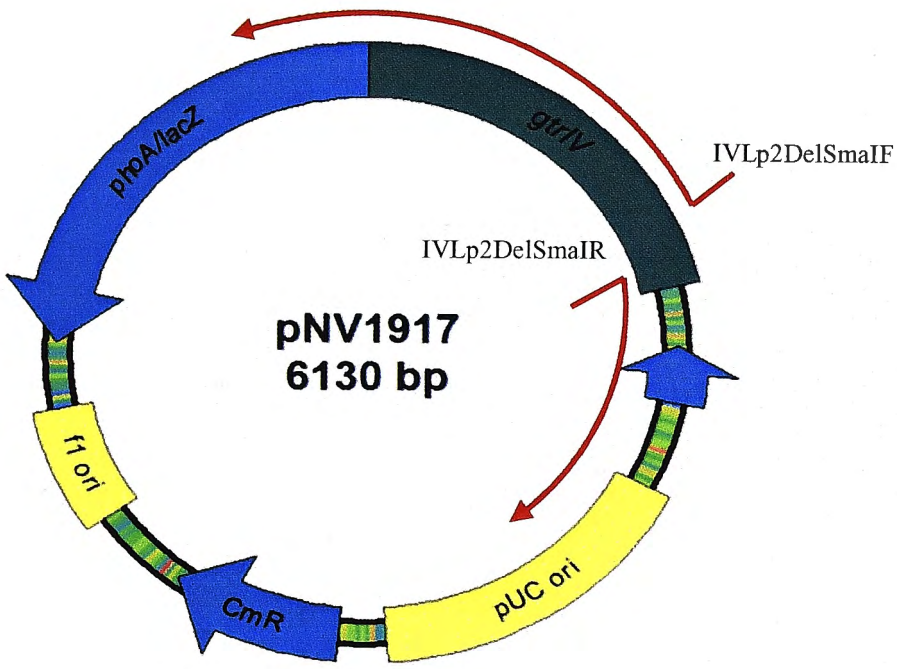


Figure 4.2: Model of GtrIV with all residues shown. Residues in pink are the 15 negatively charged amino acids present in loop No. 6. There are three negatively charged amino acid pairs DE261, ED283 and ED326. Deletion experiments for loop No. 2 involved deleting all the residues highlighted in brown, while a deletion in loop No. 3 was brought about by deleting all the residues highlighted in red. In loop No. 6, 4 different loop deletions were created. To investigate the importance of the negatively charged residues, all residues between the two black lines, which include all 15 negatively charged amino acids were deleted. To further investigate which sets of negatively charged residues were important, amino acids highlighted in green (loop No. 6 partial deletion 3), blue (loop No. 6 partial deletion 2) and orange (loop No. 6 partial deletion 1) were deleted separately.



*Dpn*I digestion
+
PCR Purification
+
*Sma*I digestion
+
Self-ligation

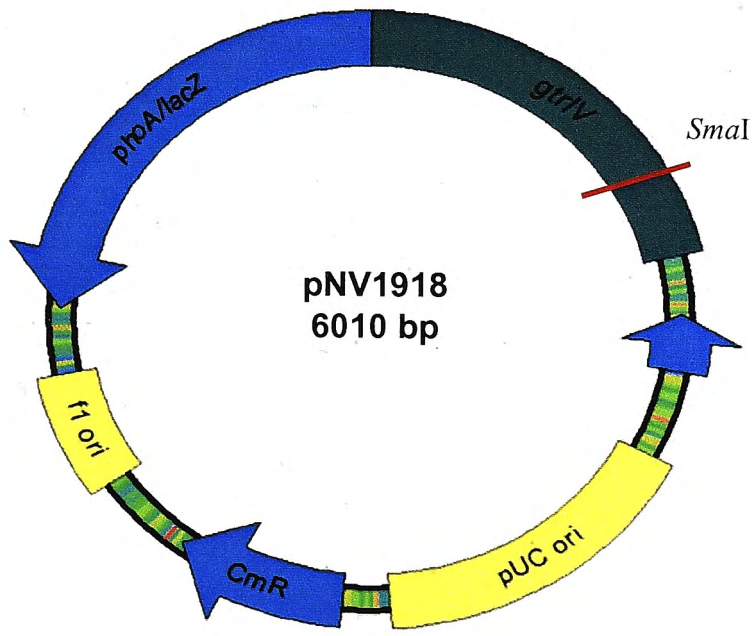
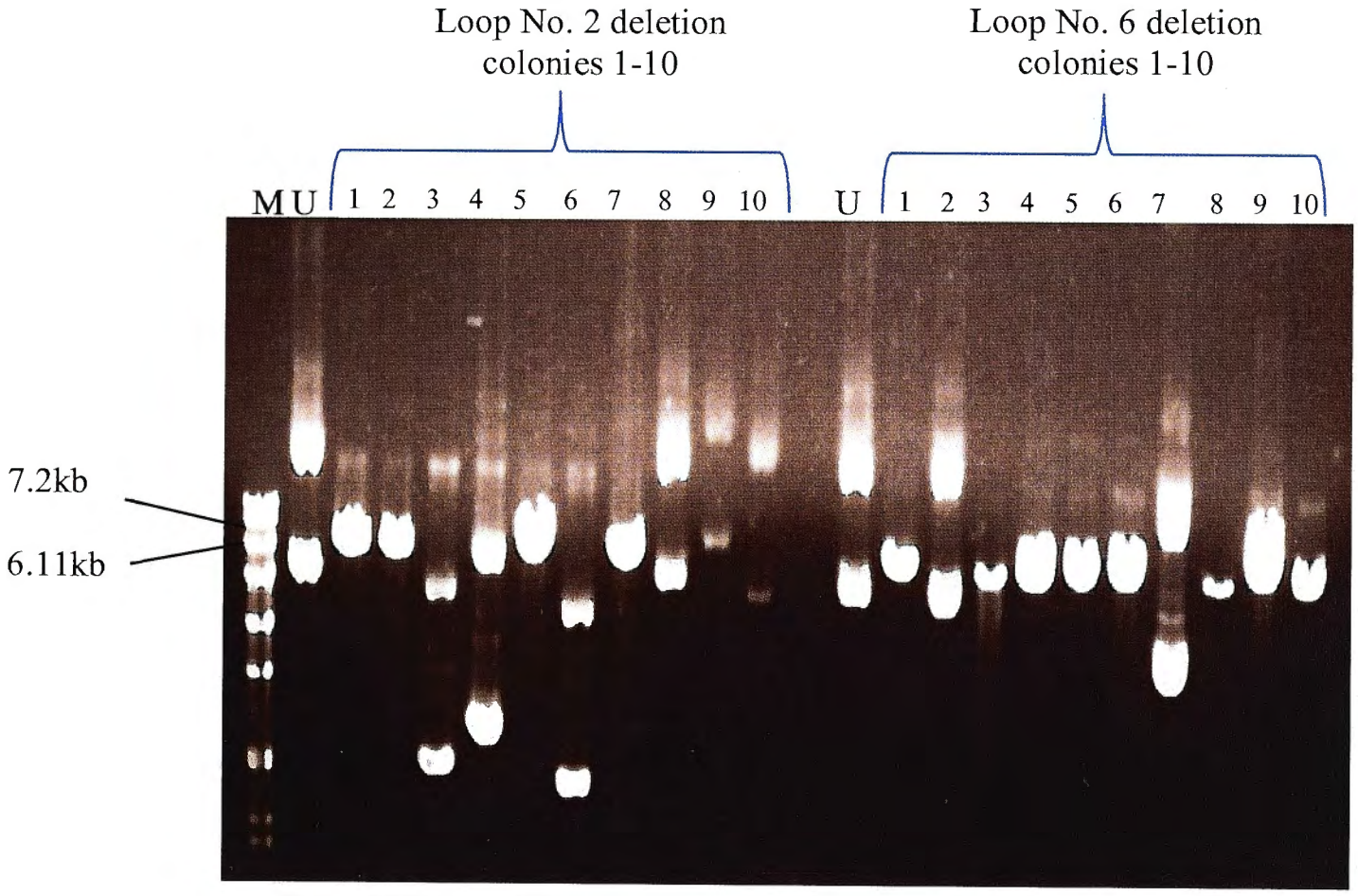


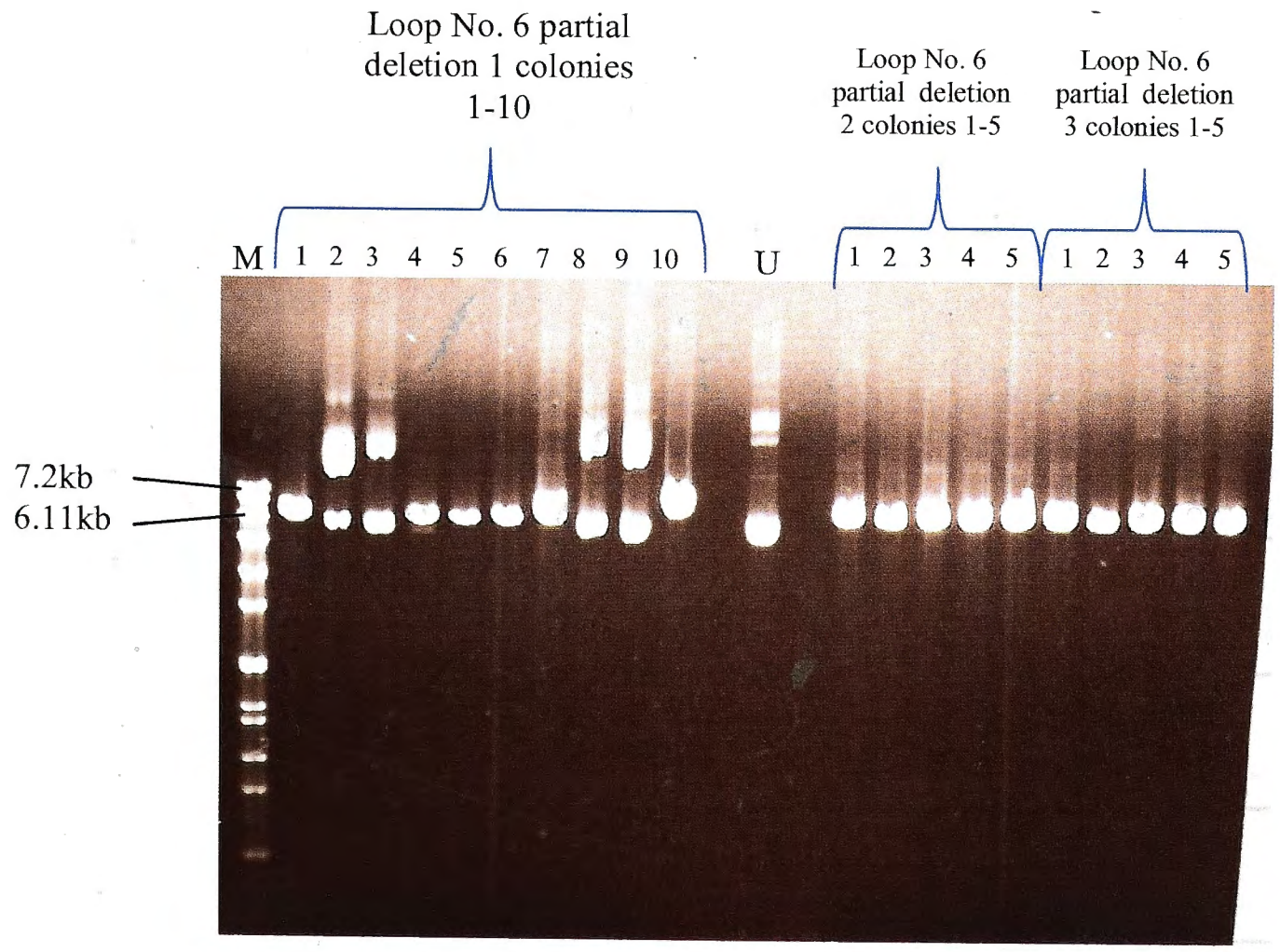
Figure 4.3: Strategy for creating loop deletion constructs. pNV1918 (GtrIV loop deletion 2) was created by amplifying most of pNV1917 with specific primers (IVLp2DelSmaIF and IVLp2DelSmaIR) such that the coding region for *gtrIV* loop No. 2 was excluded. The resulting PCR product was then *Dpn*I digested to remove traces of parental DNA present in the PCR reaction, purified and digested with *Sma*I to generate blunt ends. The *Sma*I digests were subsequently purified again to remove all traces of the restriction enzyme and the digestion buffer. The purified DNA was then self-ligated to create pNV1918. This construct now contains a *Sma*I site for screening purposes. This strategy was used to create all the loop deletion constructs in this thesis. Namely, GtrIV loop No. 3 deletion, GtrIV loop No. 6 deletion, GtrIV loop No. 6 partial deletion 1, GtrIV loop No. 6 partial deletion 2, GtrIV loop No. 6 partial deletion 3 and all resulting further deletions (FD1, 2, 3 and 4).

a)



U= Uncut pNV1917 (GtrIV fused to PhoA/LacZ

b)



U= Uncut pNV1917 (GtrIV fused to PhoA/LacZ

Figure 4.4: Screening of loop deletion clones via *Sma*I digests. a) *Sma*I digests of loop No. 2 and loop No. 6 deletion clones. 10 clones from the loop No. 2 deletion transformation plate were digested with *Sma*I. Colonies 1, 2, 5 and 7 displayed single bands that ran at the expected size of 6 kb. The rest of the colonies ran similar to the undigested DNA sample (U). Out of the 10 loop No. 6 deletion clones, 8 of them displayed the expected band size of 5.8 kb. Only colonies 2 and 7 ran similar to the uncut DNA control (U). b) *Sma*I digests of loop No. 6 partial deletions. Of the 10 loop No. 6 partial deletion 1 clones, colonies 1, 4, 5, 6 and 7 ran at the expected 6.11 kb size. All five colonies from loop No. 6 partial deletions 2 and 3 ran at the expected size of 6.11 kb. M= Molecular weight marker.

O-antigen repeat unit via an α1, 6 linkage, thus converting serotype Y to serotype 4a. Korres and Verma (2006) have suggested that the acidic residue D380 in GtrV is important for GtrV function because it is believed to interact with the O-antigen to stabilise it so that the glucosyl residue can come in contact with the correct rhamnose. Based on this hypothesis, any one or a combination of the 15 acidic residues may perform this stabilising function in GtrIV. It is also distinctly possible that these acidic residues compensate for the loss of the other acidic residues which would explain why there are three sets of paired acidic residues flanked on either side by a different acidic residue situated 7-9 residues away. To investigate this, a series of loop No. 6 deletions that targeted each set of paired acidic residues together with the flanking acidic residues were constructed (Figure 4.2). A total of 4 different loop No. 6 deletion constructs were created. These were GtrIV loop No. 6 deletion, GtrIV loop No. 6 partial deletions 1, 2 and 3. The strategy in creating these deletions was similar to that used to create pNV1918. The template used for all the deletions was pNV1917 while the primers utilised were Lp8DelSmaIF and Lp8DelSmaIR (GtrIV loop No. 6 deletion); Lp8DelSmaIF and Lp8P_Del1SmaIR (GtrIV loop No. 6 partial deletion 1); Lp8P_Del2SmaIF and Lp8P_Del2SmaIR (GtrIV loop No. 6 partial deletion 2) and finally, Lp8P_Del3SmaIF and Lp8DelSmaIR (GtrIV loop No.6 partial deletion 3). The PCR products were *Dpn*I treated, run on gel for confirmation of amplification, column purified, self-ligated and transformed into electro-competent XL1-Blue cells. Transformants from each transformation plate were screened by *Sma*I digestion (Figure 4.4a and b). Out of all the clones that were successfully digested by *Sma*I, two from each deletion was sent for sequencing to confirm for their respective deletion. The primers used to sequence each deletion construct were PhoSeqNewR and A181HpaIR. Constructs bearing the different deletions were

termed pNV1928 (Loop No. 6 deletion), pNV1929 (Loop No. 6 partial deletion 1), pNV1930 (Loop No. 6 partial deletion 2) and pNV1931 (Loop No. 6 partial deletion 3).

4.2.2 Creation of loop No. 3 deletion construct

Topology studies revealed that GtrIV also has a unique cytoplasmic loop No. 3 consisting of 30 amino acids. Considering that glucosylation occurs in the periplasm, the presence of this large cytoplasmic loop is unusual as it is not seen in the other Gtrs. To investigate its role in GtrIV function, this cytoplasmic loop was deleted. The same strategy (as above) was employed in creating this deletion construct. pNV1917 was the template used and the primers used for amplification were IVLp3DelSmaIF and IVLp3DelSmaIR. The resulting PCR product was treated with *Dpn*I, purified, digested with *Sma*I, self-ligated and transformed into electro-competent XL1-Blue cells. Resulting clones were screened via *Sma*I digestion and sent for sequencing (PhoSeqNewR and A181HpaIR primers) to confirm for the deletion. This construct was termed pNV1919.

4.2.3 Functional analysis of loop deletion constructs

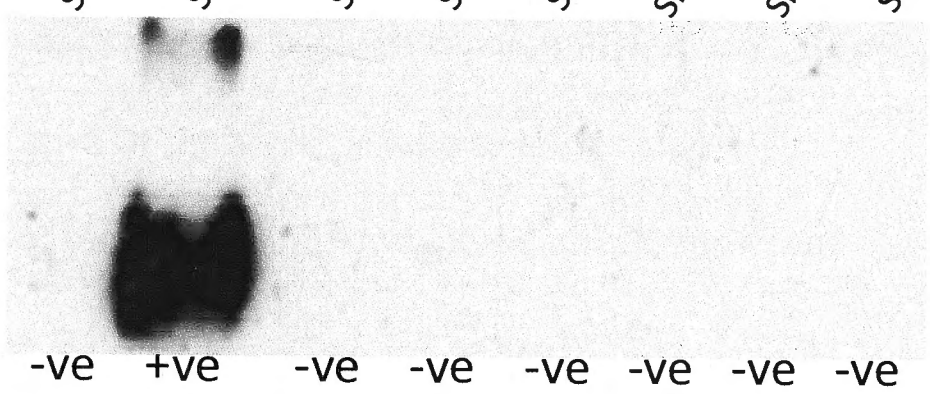
The six deletion constructs pNV1918 (loop No. 2 deletion construct), pNV1919 (loop No. 3 deletion construct), pNV1928 (loop No. 6 deletion construct), pNV1929 (loop No. 6 partial deletion 1 construct), pNV1930 (loop No. 6 partial deletion 2 construct) and pNV1931 (loop No. 6 partial deletion 3 construct) were transformed into SFL1616 (serotype Y) for functionality testing using Type IV antisera. If these GtrIV deletion proteins were to retain their function, they would convert serotype Y into serotype 4a. As such, agglutination with Type IV antisera would result. The results of all deletions were negative in a slide agglutination test using

Type IV antisera (Figure 4.5a). Although slide agglutination tests are accurate, there is a possibility that the deletions have caused a decrease in function of the protein. In such a scenario, the agglutination effect would be quite difficult to visualise. To overcome this effect, western immunoblots using Type IV antisera were performed on bacterial LPS extracted from the deletion mutants to confirm the slide agglutination results (Figure 4.5a). The results of the LPS immunoblots also showed that all the deletions were unable to convert serotype Y to serotype 4a. Deleting such large portions of a protein may have an effect on the structural stability of that protein, thus preventing it from being assembled in the bacterial cell membrane. To investigate this, western immunoblots were carried out on the membrane protein extracts of the various loop deletion constructs to determine whether the deletions had any effect on protein assembly in the membrane using anti-alkaline phosphatise antibodies (Figure 4.5b). The membrane protein western blots revealed that both loop No. 2 and loop No. 3 deletion proteins were not present on the cellular membrane. Faint banding was picked up for all the loop No. 6 deletions except for loop No. 6 partial deletion 3. The intensity of the band suggests that it is present in almost the same quantity as the undeleted GtrIV. This suggests that the 19 amino acids that were deleted could play a part in the enzyme's catalytic function rather than marinating the proteins structural integrity.

4.3 GtrIV loop No. 6 further deletions

Based on the loop deletion results, four new deletion constructs that specifically targeted the 19 amino acid spanning region that made up loop No. 6 partial deletion 3 were created (Figure 4.6a). These further deletions were termed further deletions (FD) 1, 2, 3 and 4. The strategy used to create these new deletion constructs was identical to the one used in creating the

a)


SFL 1616 (Serotype Y, negative control)
SFL 2324 (Serotype 4a, Positive control)
SFL 2325 (Loop No. 2 deletion)
SFL 2326 (Loop No. 3 deletion)
SFL 2328 (Loop No. 6 deletion)
SFL 2327 (Loop No. 6 partial deletion 1)
SFL 2330 (Loop No. 6 partial deletion 2)
SFL 2329 (Loop No. 6 partial deletion 3)
-ve
+ve
-ve
-ve
-ve
-ve
-ve
-ve

b)

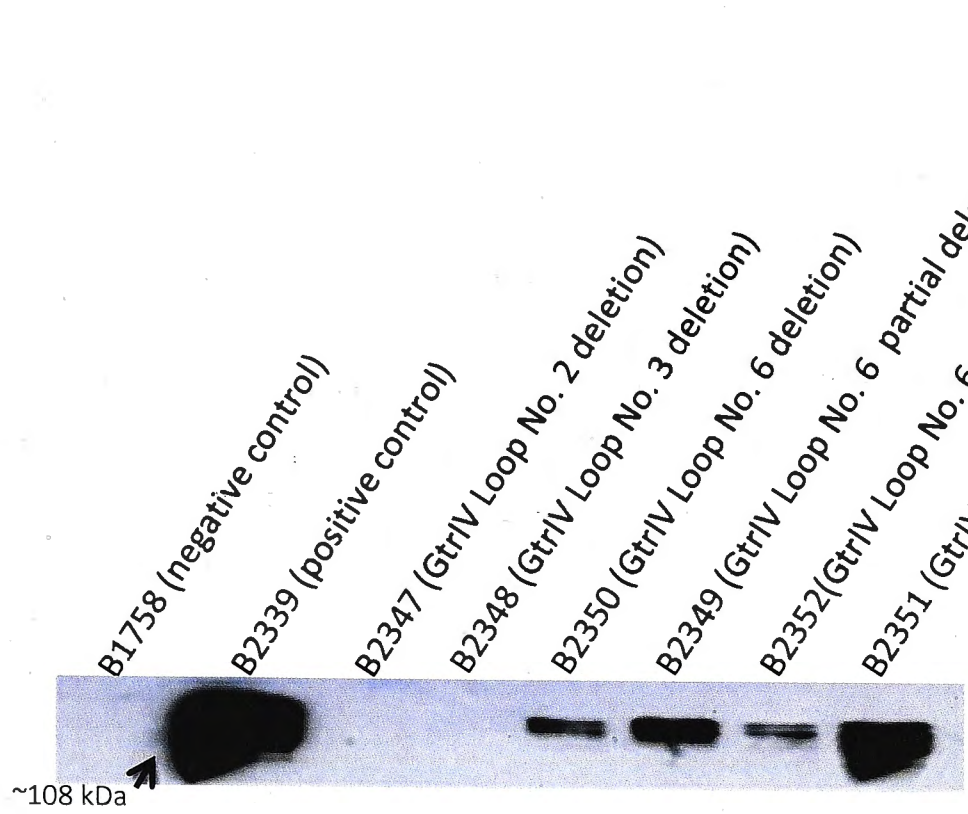

B1758 (negative control)
B2339 (positive control)
B2347 (GtrIV Loop No. 2 deletion)
B2348 (GtrIV Loop No. 3 deletion)
B2350 (GtrIV Loop No. 6 deletion)
B2349 (GtrIV Loop No. 6 partial deletion 1)
B2352(GtrIV Loop No. 6 partial deletion 2)
B2351 (GtrIV Loop No. 6 partial deletion 3)
~108 kDa

Figure 4.5: Functional analysis of deletion constructs. a) LPS Western blots and slide agglutination results. Deletion constructs were transformed into a serotype Y strain (SFL1616) to test for modification of O-antigen to serotype 4a. SFL 2324 is GtrIV fused to PhoA/LacZ and then transformed into SFL1616. The bright band is indicative of the serotype conversion from serotype Y to serotype 4a. The rest of the deletions show absence of modification to serotype 4a when probed with Type IV antisera. Results of the slide agglutination assays are indicated below the Western blots. A score of +ve indicates a positive result while a score of –ve indicates a negative result. b) Western blots (using anti-alkaline phosphatase primary antibody) performed on the membrane protein extractions of the various GtrIV loop deletion constructs that were fused to alkaline phosphatase. B2347 and B2348 were not detected, indicating GtrIV Loop No. 2 and Loop No. 3 deletions have a direct effect on the assembly of the protein in the cellular membrane. B2349, B2350 and B2352 were detected in lower amounts, the amount of protein that is assembled in the cellular membrane in these three deletion mutants is much lower as compared to the fully intact protein (B2339). The intense band seen for B2351 is almost the same as the fully intact protein.

a)

GtrIV : NGKENPSWAVDECIGVDAW
251 269

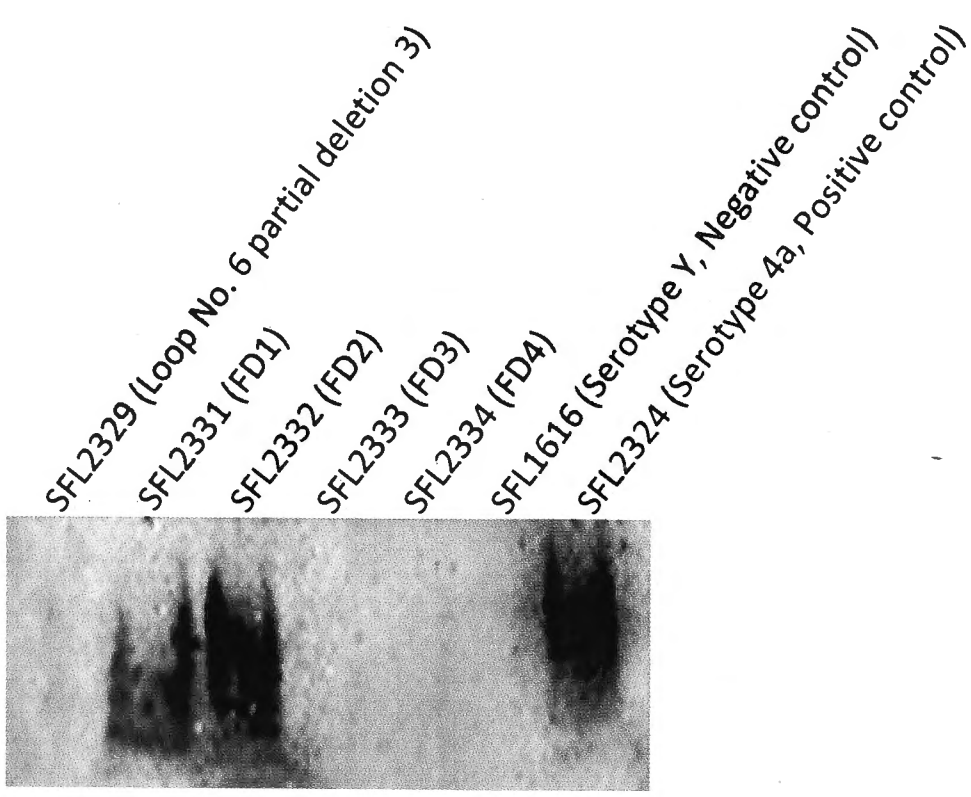
GtrIV FD1 : -----PSWAVDECIGVDAW
251 269

GtrIV FD2 : NGKEN-----DECIGVDAW
251 269

GtrIV FD3 : NGKENPSWAV-----VDAW
251 269

GtrIV FD4 : NGKENPSWAVDECIG----
251 269

b)



c)

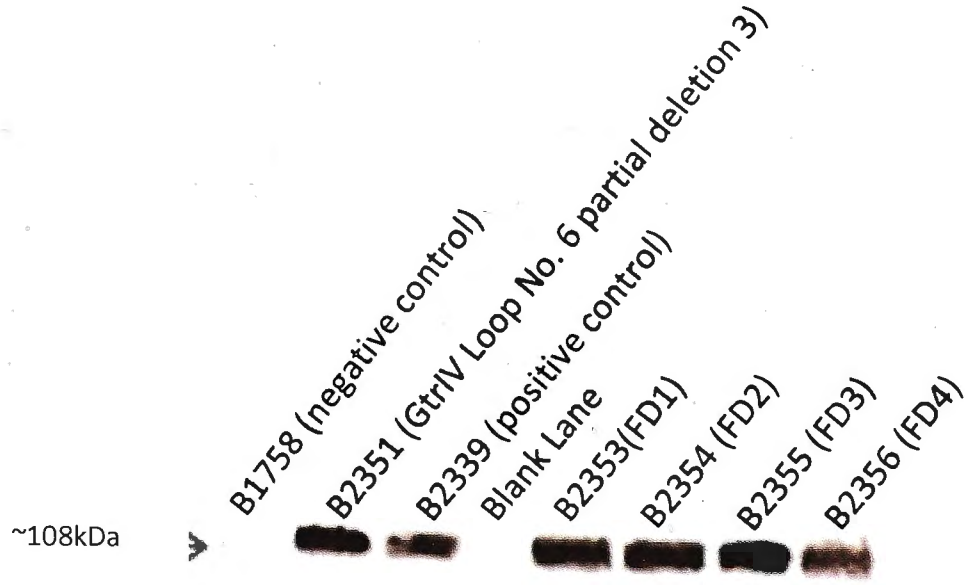


Figure 4.6: Further deletions of GtrIV loop No. 6. a) The sequence of the 19 amino acids located in GtrIV loop No. 6, which were targeted for further deletion. The dashes in the sequences of GtrIV FD1, FD2, FD3 and FD4 indicate the amino acids that were deleted in each deletion set. Amino acid numbers corresponding to the start and end of the peptide sequence are given below their respective residue. b) LPS Western blots of GtrIV deletion constructs that were transformed into SFL1616 (Serotype Y). Deletion constructs FD1(SFL2331) and FD2 (SFL2332) were able to convert serotype Y to serotype 4a, thus indicating functionality. The absence of bands for FD3 and FD4 indicates loss of function for these deletion mutants. c) Western blots of GtrIV Loop No. 6 further deletion constructs. All the further deletion constructs were detected when the membrane extracts were probed with anti-alkaline phosphatase. This indicated that the deletion proteins (which are fused to PhoA/LacZ) are expressed.

previous loop No. 6 deletion constructs. Again, the template used for the PCR reactions was pNV1917. The primers used were IVLp6P.Del3FD1F and Lp8DelSmaIR (FD1); IVLp6P.Del3FD2F and IVLp6P.Del3FD2R (FD2); IVLp6P.Del3FD3F and IVLp6P.Del3FD3R (FD3) and finally, Lp8DelSmaIF and IVLp6P.Del3FD4R (FD4). Resulting clones were screened via *Sma*I digests and confirmed by sequencing. A total of 4 constructs (FD1, 2, 3 and 4) were obtained and used to transform SFL1616. Functional analysis of GtrIV deletion constructs through slide agglutination and LPS Western blots revealed that the introduced deletions FD1 and FD2 were tolerated whereas the deletions present in FD3 and FD4 knocked out GtrIV function (Figure 4.6b). Western blots of the membrane proteins extracted from these constructs show that all the four further deletion proteins were localised in the membrane (Figure 4.6c).

4.4 Investigating periplasmic loop function of GtrIV by using chimeric proteins

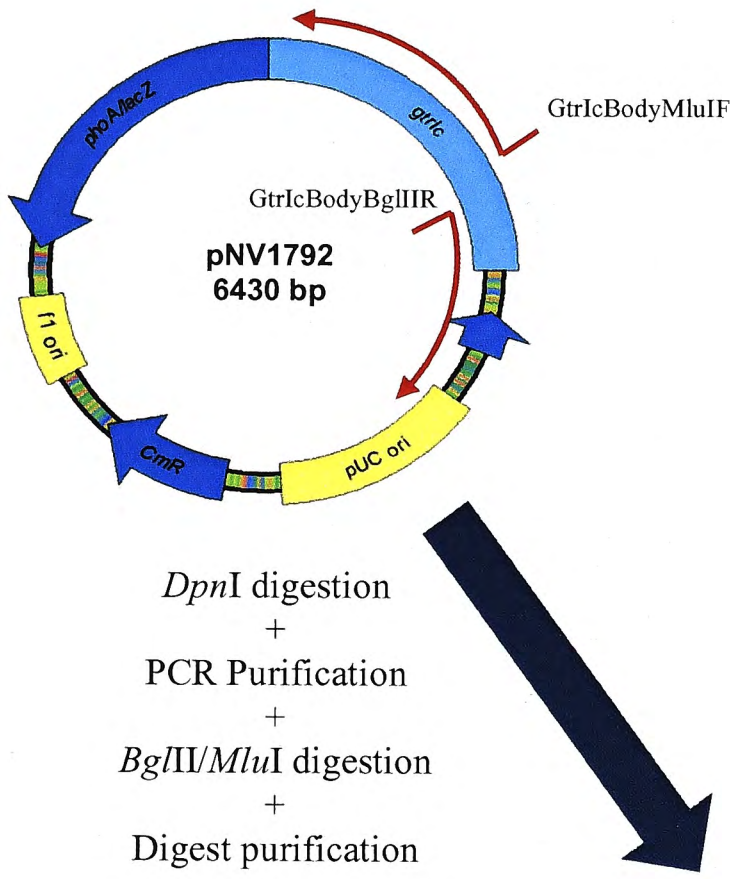
4.4.1 Creation of GtrIc and GtrIV loop No. 2 chimeras

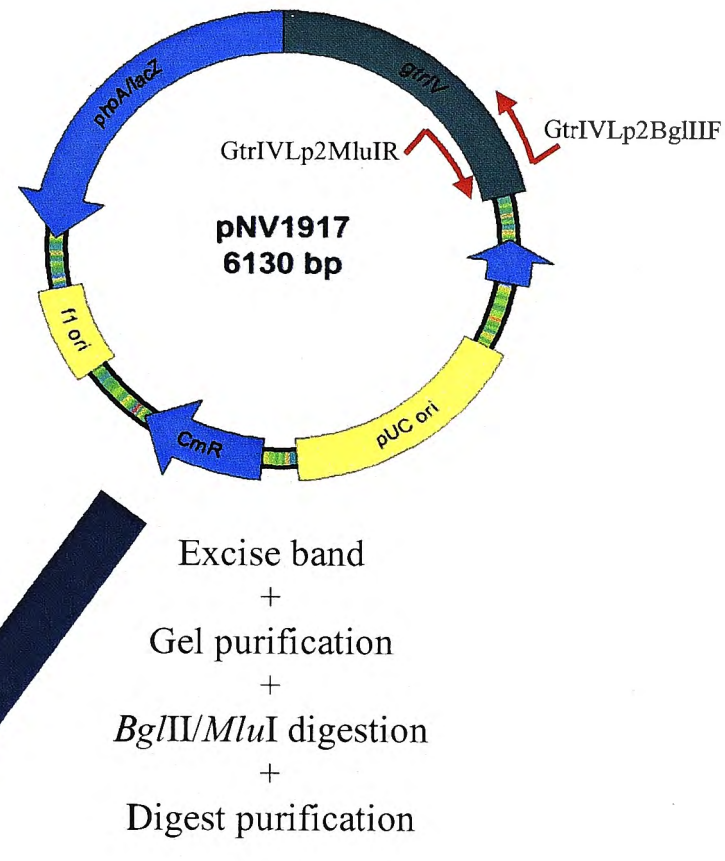
As previous studies on other $Gtr_{(type)}$ proteins have shown that a degree of functional similarity exists between the $Gtr_{(type)}$ family , the role of loop No. 2 of GtrIV by creating chimeric proteins with its closest structural homologue GtrIc was investigated. The accepted dogma is that topologically similar N-terminal periplasmic loops in all $Gtr_{(type)}$ proteins may perform the conserved functions of interacting with the UndP-glucose and possibly recycling it to the cytoplasmic face. Thus, if both the N-terminal periplasmic loops No. 2 of GtrIV and GtrIc play the same role in O-antigen modification, then swapping these two loops with each other should not abolish function of the native proteins. Via PCR amplification, the sequences encoding the

gtrIc and *gtrIV* loop No. 2 regions were amplified. pNV1792 (carries *gtrIc* in tandem with *phoA/lacZ* in pBC SK) and pNV1917 (carries *gtrIV* in tandem with *phoA/lacZ* in pBC SK) were also amplified by PCR such that the sequences that encode for the loop No. 2 regions were excluded. A detailed schematic of this strategy is presented in Figures 4.7 and 4.8. After transformation into electro-competent XL1-Blue cells, transformants from each plate were screened via a series of *Bgl*II and *Mlu*I restriction digests (Figure 4.9a and b). Successful digestion by both enzymes indicated that the inserts were ligated with their respective vectors in the correct orientation, thus regenerating the restriction sites. The positive clones were sequenced to confirm for correct insertion of the loop No. 2 coding sequences to their respective vectors. The GtrIV-GtrIc loop No.2-GtrIV construct was labelled pNV1937 while the GtrIc-GtrIV loop No.2-GtrIc chimeric construct was labelled pNV1936. Table 4.1 shows what GtrIc and GtrIV models would look like with their loop No 2 exchanged.

4.4.2 Creation of GtrIc-GtrIV loop No. 6-GtrIc and GtrIV-GtrIc loop No. 10-GtrIV chimeras

In parallel to investigating the role of GtrIV loop No. 2, we also investigated the role of the C-terminal periplasmic loop No. 6 of GtrIV. Gtrs I, II, V and X have long periplasmic C-terminal ends that have been hypothesised to be responsible for conferring serotype specificity. In contrast, as both GtrIV and GtrIc have short cytoplasmic C-terminal ends, their specific function is thought to be performed by the large periplasmic loops No. 6 and No. 10 for GtrIV and GtrIc, respectively. By using the same PCR-based approach for the creation of the loop No. 2 chimeras, loop No. 6 of GtrIV was replaced by loop No. 10 of GtrIc and vice versa. Figure 4.10 details the creation of pNV1938 (GtrIc-GtrIV loop No. 6-GtrIc chimera). The primers used to amplify pNV1792 introduced *Sma*I sites while the primers used to amplify *gtrIV* loop No. 6





*Dpn*I digestion
+
PCR Purification
+
*Bgl*II/*Mlu*I digestion
+
Digest purification

Excise band
+
Gel purification
+
*Bgl*II/*Mlu*I digestion
+
Digest purification

Ligation

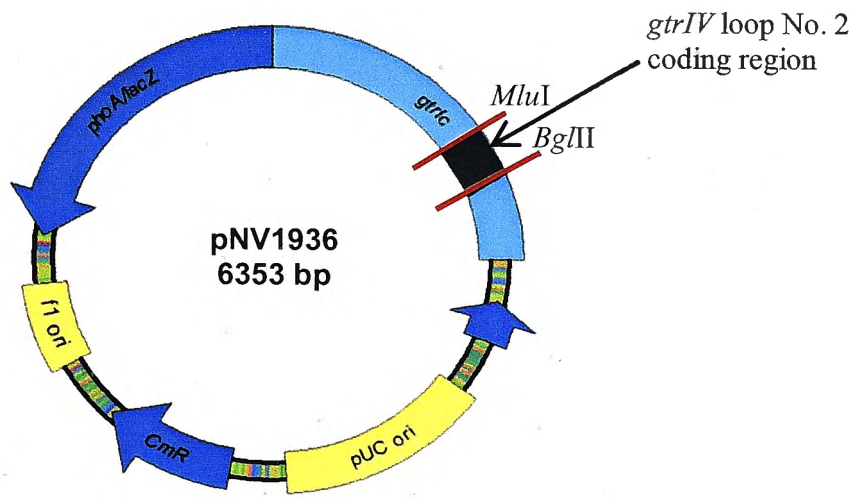
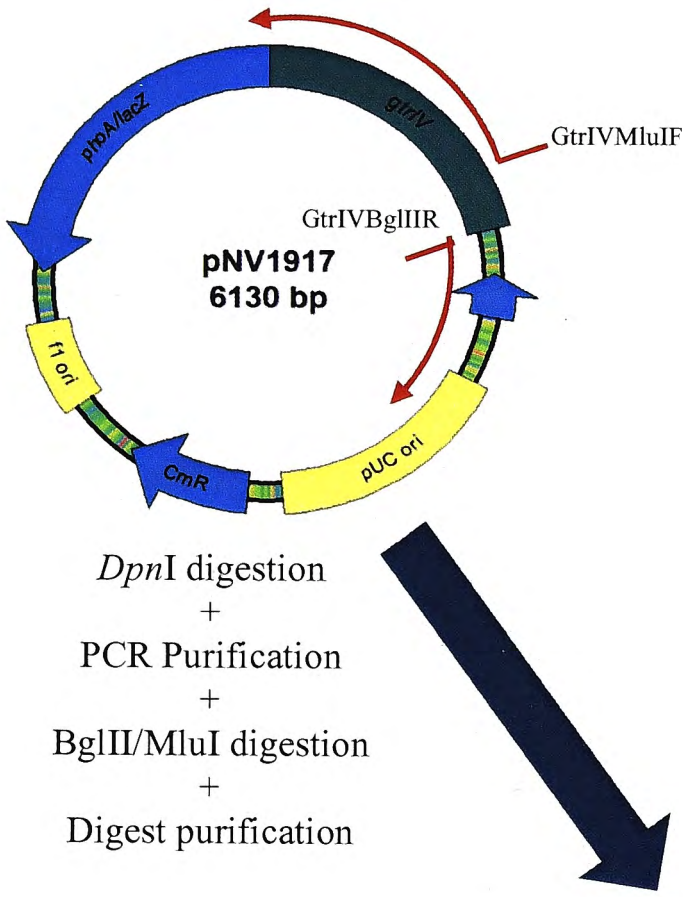
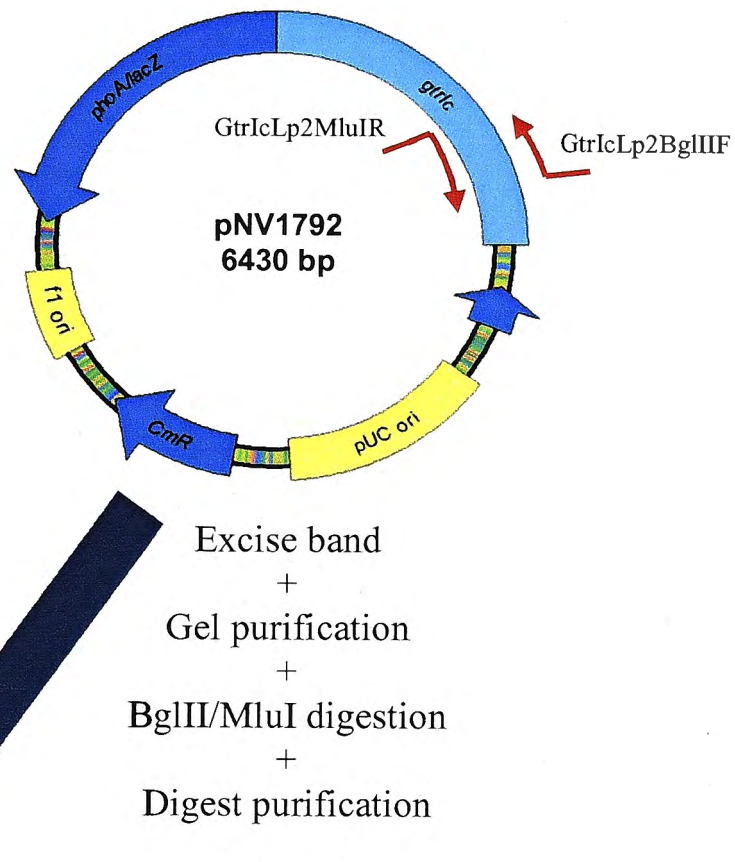


Figure 4.7: Strategy for creating pNV1936 (GtrIc-GtrIV loop No. 2-GtrIc chimera). The primers GtrIcBodyMluIF and GtrIcBodyBglIIR were used to amplify most of pNV1792 excluding the coding region of *gtrIc* loop No. 2. The amplified DNA was then treated with *Dpn*I to remove traces of parental DNA and purified. It was then digested with *Bgl*II and *Mlu*I to generate sticky ends for ligation. In parallel, the coding region for *gtrIV* loop No. 2 was amplified from pNV1917 with the help of IVLp2BglIIF and IVLp2MluIR, run on 1.5% gel and the corresponding gel band was excised and purified. This purified product was subsequently digested with *Bgl*II and *Mlu*I to generate compatible sticky ends for ligation with the amplified pNV1792. The resulting product was termed pNV1936. This hybrid construct contains most of pNV1792, as well as the region encoding *gtrIV* loop No. 2. The insert is flanked by *Bgl*II and *Mlu*I restriction sites for screening purposes.



*Dpn*I digestion
+
PCR Purification
+
BglII/MluI digestion
+
Digest purification



Excise band
+
Gel purification
+
BglII/MluI digestion
+
Digest purification

Ligation

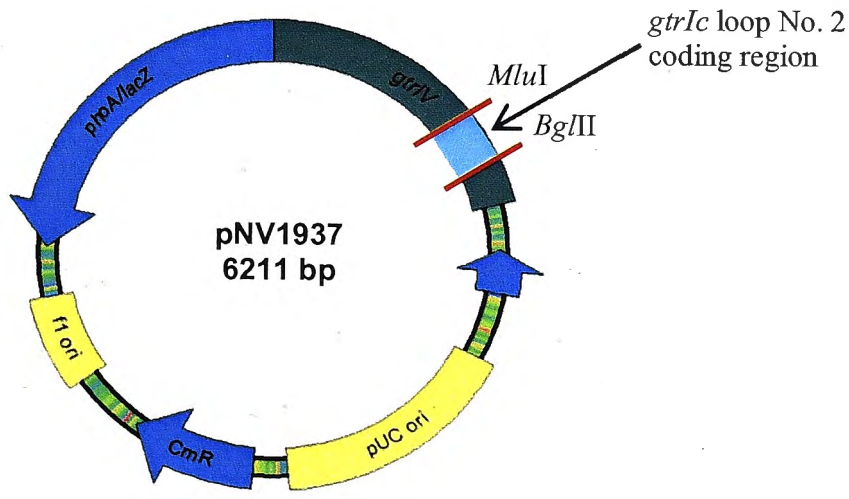
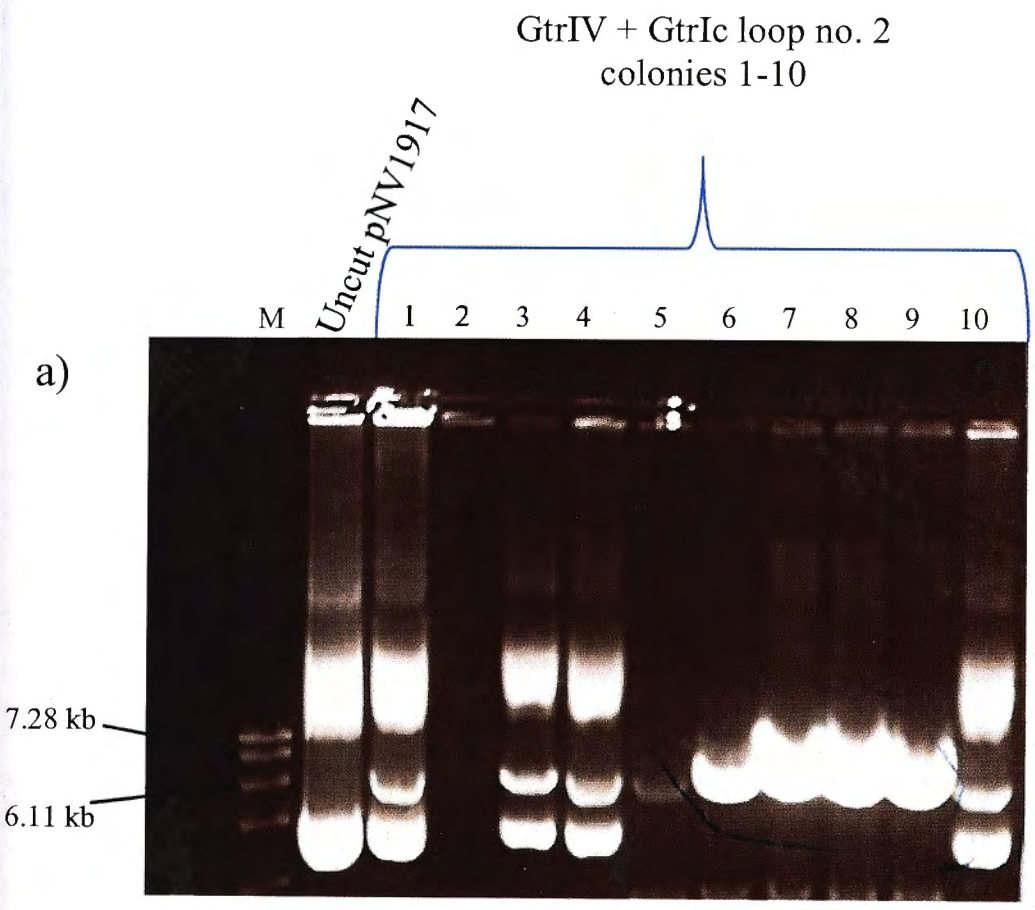
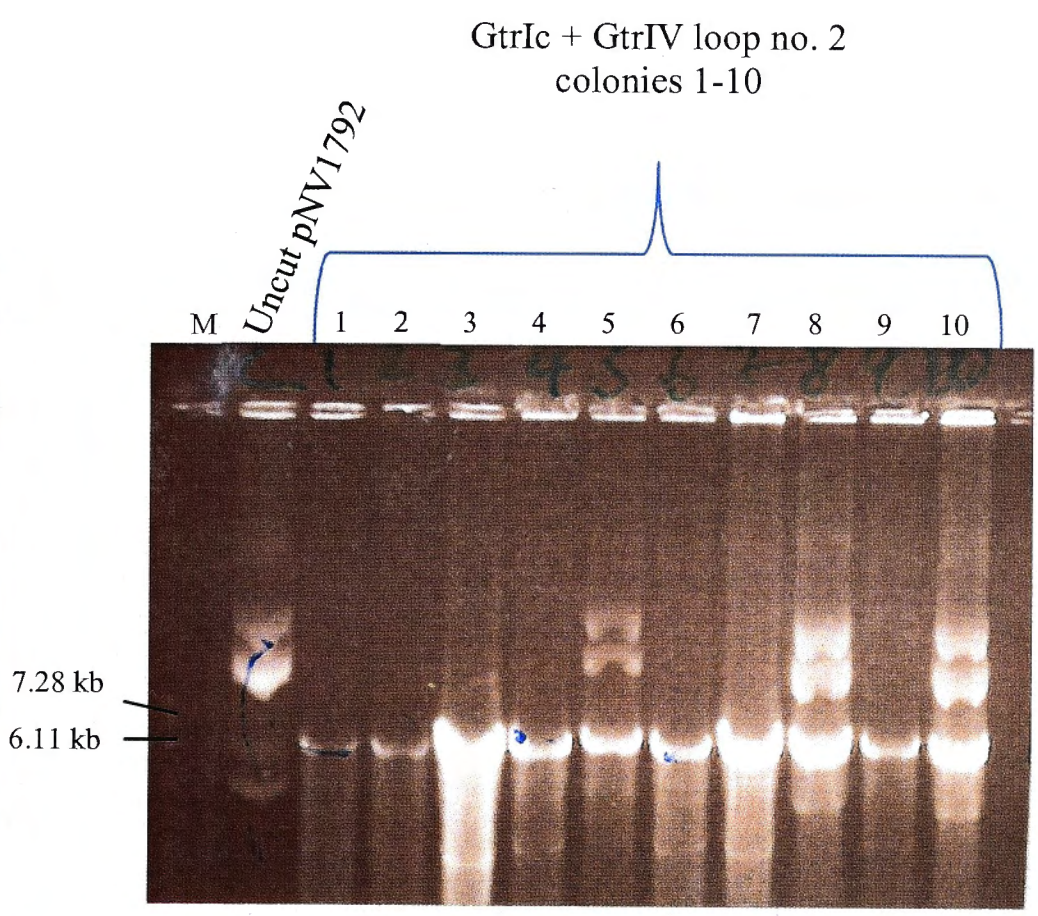


Figure 4.8: Strategy for creating pNV1937 (GtrIV-GtrIc loop No. 2-GtrIV chimera). Most of pNV1917 was amplified with the primers GtrIVMluIF and GtrIVBglIIR such that the coding region of *gtrIV* loop No. 2 was excluded. Traces of parental DNA from the amplified DNA were removed by treatment with *Dpn*I. After DNA purification, sticky ends were generated by digestion with *Bgl*II and *Mlu*I. At the same time, the coding region for *gtrIc* loop No. 2 was amplified from pNV1792 with the help of GtrIcLp2BglIIF and GtrIcLp2MluIR, run on 1.5% gel and the corresponding gel band was excised and purified. Compatible sticky ends were generated digesting the amplified DNA with *Bgl*II and *Mlu*I. The resulting product is pNV1937. This plasmid contains most of pNV1917, as well as the region encoding *gtrIc* loop No. 2. The insert is flanked by *Bgl*II and *Mlu*I restriction sites for screening purposes.

a)


GtrIV + GtrIc loop no. 2
colonies 1-10
Uncut pNV1917
M
1 2 3 4 5 6 7 8 9 10
7.28 kb
6.11 kb


GtrIc + GtrIV loop no. 2
colonies 1-10
Uncut pNV1792
M
1 2 3 4 5 6 7 8 9 10
7.28 kb
6.11 kb

b)

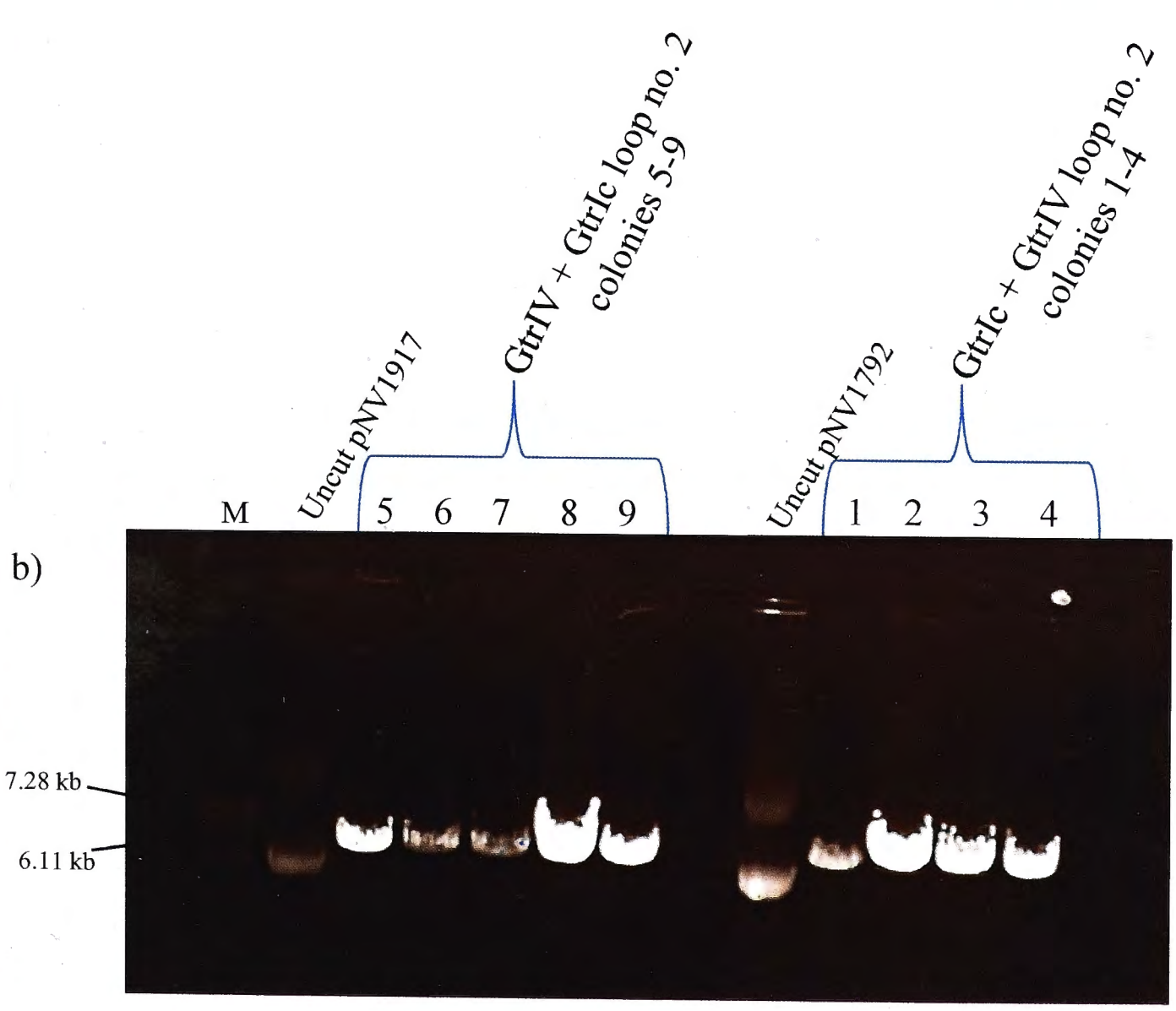
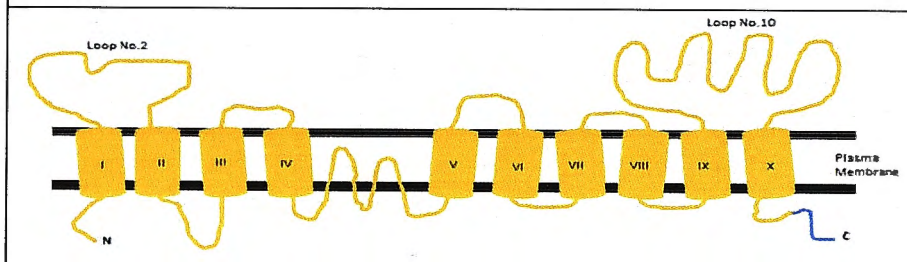
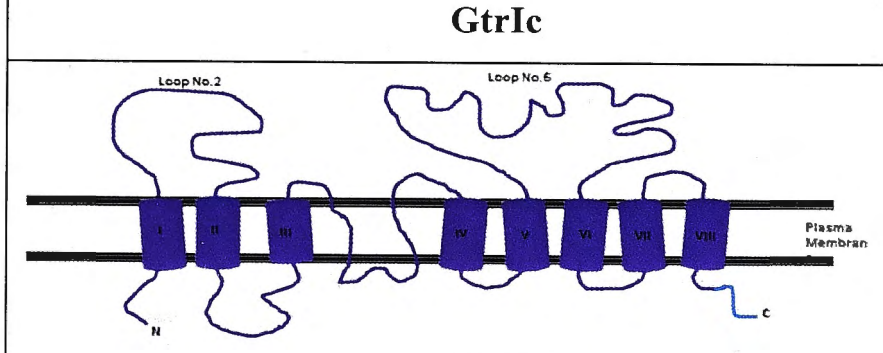
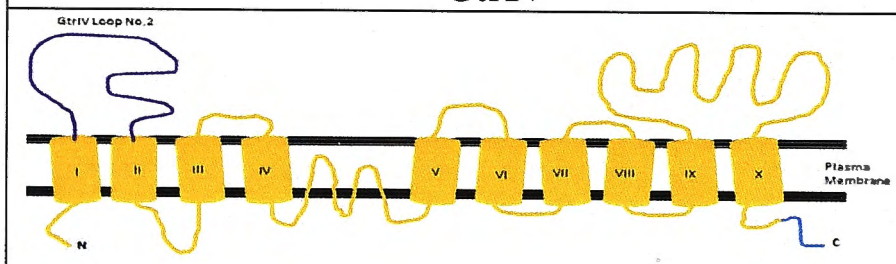
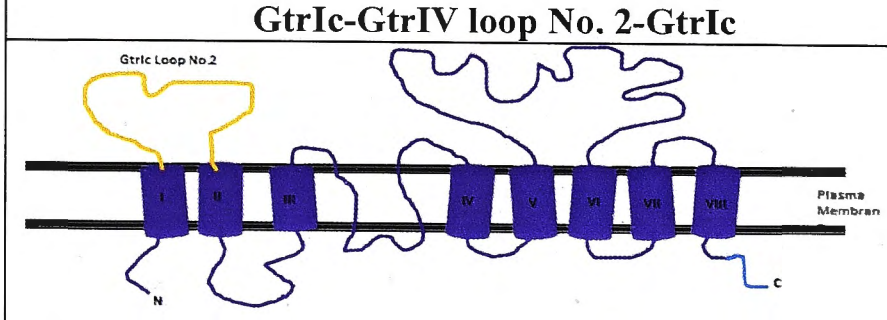
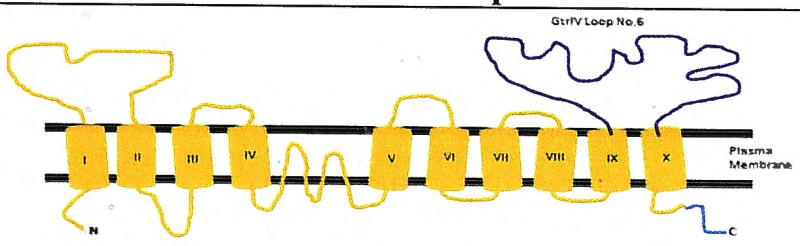
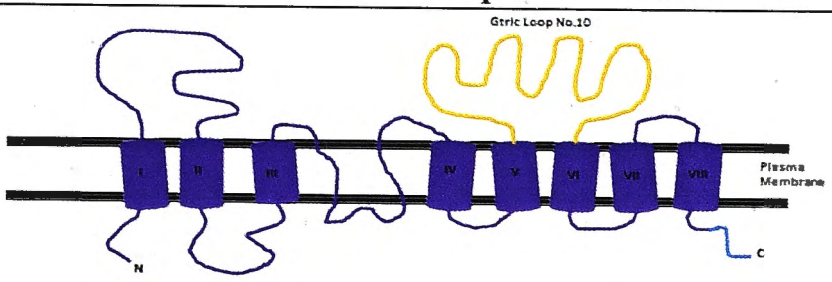
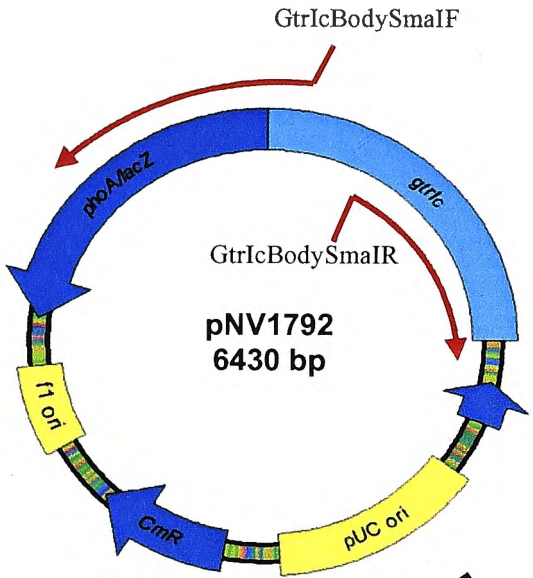

GtrIV + GtrIc loop no. 2
colonies 5-9
Uncut pNV1917
GtrIc + GtrIV loop no. 2
colonies 1-4
Uncut pNV1792
M
5 6 7 8 9
1 2 3 4
7.28 kb
6.11 kb

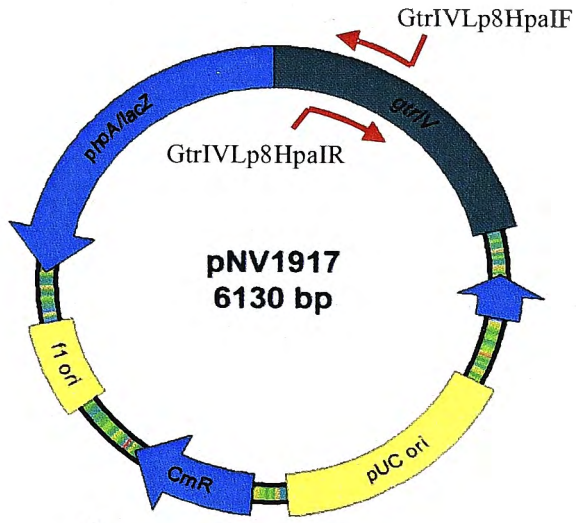
Figure 4.9: Screening of GtrIV+GtrIc loop No.2 and GtrIc+GtrIV loop No. 2 clones via *Bgl*II and *Mlu*I restriction digests. a) 10 transformants from the GtrIV+GtrIc loop No. 2 transformation plates were subject to *Bgl*II digests. Colonies 5-9 displayed bands at the expected 6.2kb region and were deemed to be successfully digested with *Bgl*II. 10 transformants from the GtrIc+GtrIV loop No. 2 plates were also digested with *Bgl*II. Colonies 1, 2, 3, 4, 6, 7 and 9 produced the expected 6.3 kb bands and were deemed to be digested with *Bgl*II. b) GtrIV+GtrIc loop No. 2 colonies 5-9 and GtrIc+GtrIV loop No. 2 colonies 1-4 were digested with *Mlu*I for further confirmation. The resulting bands from GtrIV+GtrIc loop No. 2 colonies 5-9 ran at 6.2 kb while the bands of GtrIc+GtrIV loop No. 2 colonies 1-4 ran at ~6.3 kb. These digests confirm that GtrIV loop No. 2 and GtrIc loop No. 2 were inserted in the right orientation. M= Molecular weight marker.

Table 4.1: Predicted structures of the four chimeric proteins created in this study between GtrIc and GtrIV.

Chimera Constructs	***Slide Agglutination***	
	MASF Ic	Type IV
 **GtrIc**	+ve	-ve
 **GtrIV**	-ve	+ve
 **GtrIc-GtrIV loop No. 2-GtrIc**	-ve	-ve
 **GtrIV-GtrIc loop No. 2-GtrIV**	-ve	-ve
 **GtrIc-GtrIV loop No.6-GtrIc**	-ve	-ve
 **GtrIV-GtrIc loop No.10-GtrIV**	-ve	-ve

Points of fusion : a) GtrIc-GtrIV loop No.2-GtrIc chimera :GtrIc-F31/GtrIV-N27, GtrIV-Y68/GtrIc-V100, b) GtrIV-GtrIc loop No.2-GtrIV chimera: GtrIV-M26/GtrIc-H32, GtrIc-N99/GtrIV-S71, c) GtrIc-GtrIV loop No.6-GtrIc chimera: GtrIc-P382/GtrIV-T231, GtrIV-K351/GtrIc-V495, d) GtrIV-GtrIc loop No.10-GtrIV chimera: GtrIV-S229/GtrIc-T231, GtrIc-V381/GtrIV-D352. To test for GtrIc function, the chimera constructs were transformed in to SFL1416 (Serotype 1a) to test for serotype conversion to serotype 1c using MASF Ic. As for GtrIV function, conversion to serotype 4a was tested by using Type IV antisera against constructs that were transformed into SFL1616 (serotype Y).





*Dpn*I digestion
+
PCR Purification
+
*Sma*I digestion
+
Digest purification

Excise band
+
Gel purification
+
*Hpa*I digestion
+
Digest purification

Ligation

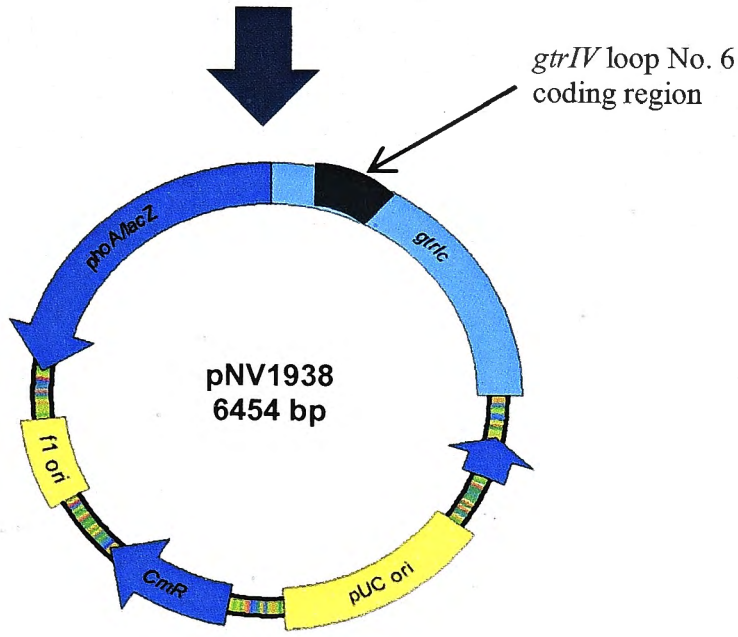


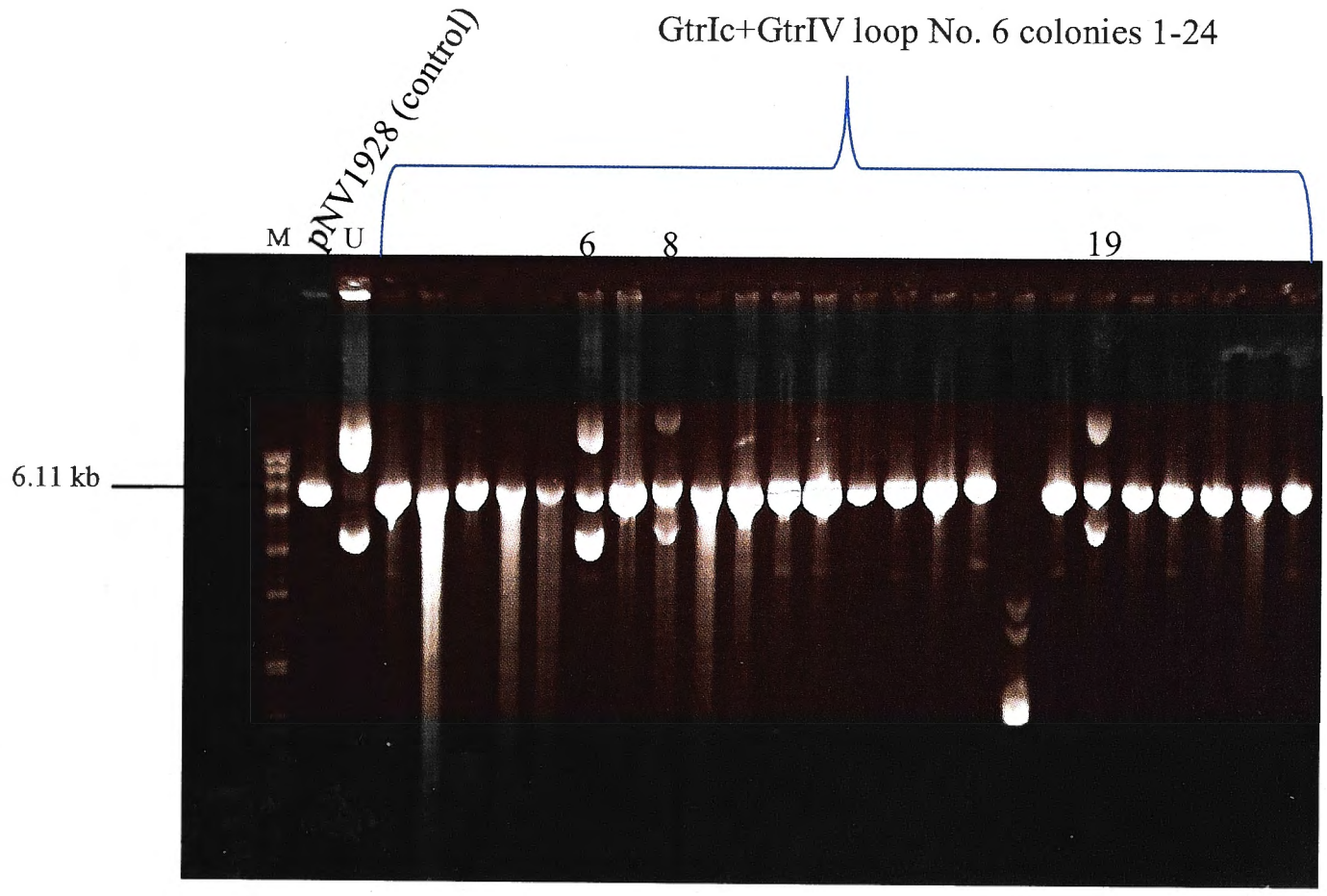
Figure 4.10: Creation of pNV1938 (GtrIc-GtrIV loop No. 6-GtrIc chimera). The majority of pNV1792 was amplified using the primers GtrIcBodySmaIF and GtrIcBodySmaIR to exclude the *gtrIc* loop no. 2 coding region. The PCR product was treated with *Dpn*I, purified and digested with *Sma*I to generate blunt ends for ligation. The coding region for *gtrIV* loop No. 6 was amplified from pNV1917 using GtrIVLp8HpaIF and GtrIVLp8HpaIR. The PCR product was run on 1.5% agarose gel and the corresponding band was excised. The gel fragment was then purified and digested with *Hpa*I to generate blunt ends for ligation. Both sets of DNA were ligated together to create pNV1938. This plasmid contains most of pNV1792, as well as the region encoding *gtrIV* loop No. 6. As two different restriction sites were used in the PCR amplification, both sites are destroyed upon ligation. As a result, the insert is not flanked by any restriction sites.

introduced *Hpa*I sites. Following amplification by PCR of these sequences, restriction digests with *Sma*I (amplified *gtrIc*) and *Hpa*I (amplified *gtrIV* loop No. 6 coding region) was carried out followed by ligation of *gtrIV* loop No. 6 with *gtrIc* and transformation of the ligation mix into XL1-Blue cells. Because the restriction sites were dissimilar, recombinant plasmids carrying the correct chimeric construct were therefore not expected to contain a *Sma*I restriction site. As pNV1792 already contains a *Hpa*I site, the resulting transformants were screened via *Sma*I digestion (Figure 4.11). All clones that were not digested with *Sma*I were selected and sequenced to confirm that the *gtrIV* loop No. 6 coding region was inserted in- frame within *gtrIc*.

To create pNV1939 ((GtrIV-GtrIc loop No. 10-GtrIV chimera) (detailed in Figure 4.12), pNV1917 was amplified with the primers GtrIVBodySmaIF and GtrIVBodySmaIR to exclude the *gtrIV* loop No. 6 coding region. The *gtrIc* loop No. 10 coding region was amplified from pNV1792 with the help of the primers GtrIcLp10SmaIF and GtrIcLp10SmaIR. All four primers incorporated *Sma*I sites into the amplified DNA. After transformation of the ligation mix into electro-competent XL1-Blue cells, recombinant plasmids carrying this chimeric construct were therefore expected to have two regenerated *Sma*I sites flanking the *gtrIc* loop No. 10 sequence within *gtrIV*. All resulting transformants were screened via *Sma*I digestion (Figure 4.13). Clones that exhibited two bands (~6.1kb and 0.3kb) were sent for sequencing to confirm that *gtrIc* loop No. 10 was successfully inserted in the correct orientation.

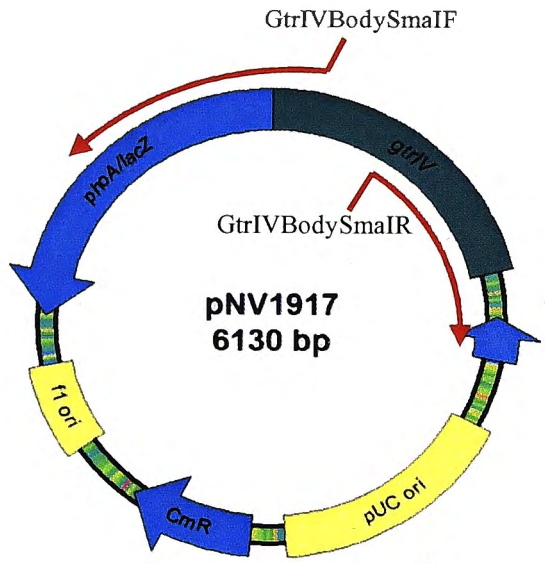
4.4.3 Functional analysis of GtrIc and GtrIV chimeric proteins

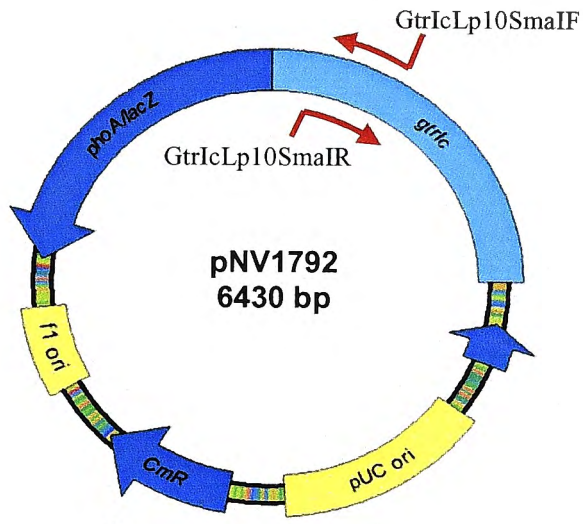
Functional analysis of the various GtrIV and GtrIc chimeric proteins was carried out via slide agglutination using Type IV antisera and MASF Ic (monoclonal antibodies specific to



U = Uncut pNV1928 (GtrIV loop No. 6 deletion)

Figure 4.11: Screening of GtrIc-GtrIV loop No. 6-GtrIc chimeras via *Sma*I digestion. 24 colonies from the GtrIc-GtrIV loop No. 6-GtrIc transformation plate were digested with *Sma*I. Colonies 6, 8 and 19 run the same as the uncut control (U). This indicates that they were not digested by *Sma*I. These colonies also ran differently from the *Sma*I digested control (C) pNV1928 (GtrIV loop No. 6 deletion construct). M= Molecular weight marker





*Dpn*I digestion
+
PCR Purification
+
*Sma*I digestion
+
Digest purification

Excise band
+
Gel purification
+
*Sma*I digestion
+
Digest purification

Ligation

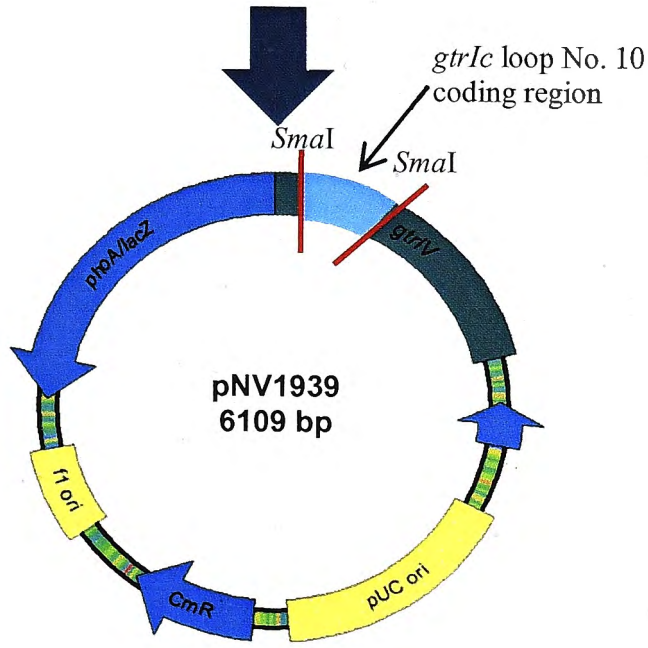


Figure 4.12: Creation of pNV1939 (GtrIV-GtrIc loop No. 10-GtrIV chimera). The majority of pNV1917 was amplified using the primers GtrIVBodySmaIF and GtrIVBodySmaIR to exclude the *gtrIV* loop no. 2 coding region. The PCR product was *Dpn*I treated, purified and digested with *Sma*I to generate blunt ends for ligation. In parallel, the coding region for *gtrIc* loop No. 10 was amplified from pNV1792. The resulting PCR product was run on 1.5% agarose gel and the corresponding band was excised. Upon purification and *Sma*I digestion, both sets of DNA were ligated together to create pNV1939. This hybrid construct contains most of pNV1792, as well as the region encoding *gtrIc* loop No. 10. The insert is flanked by *Sma*I sites for screening purposes.

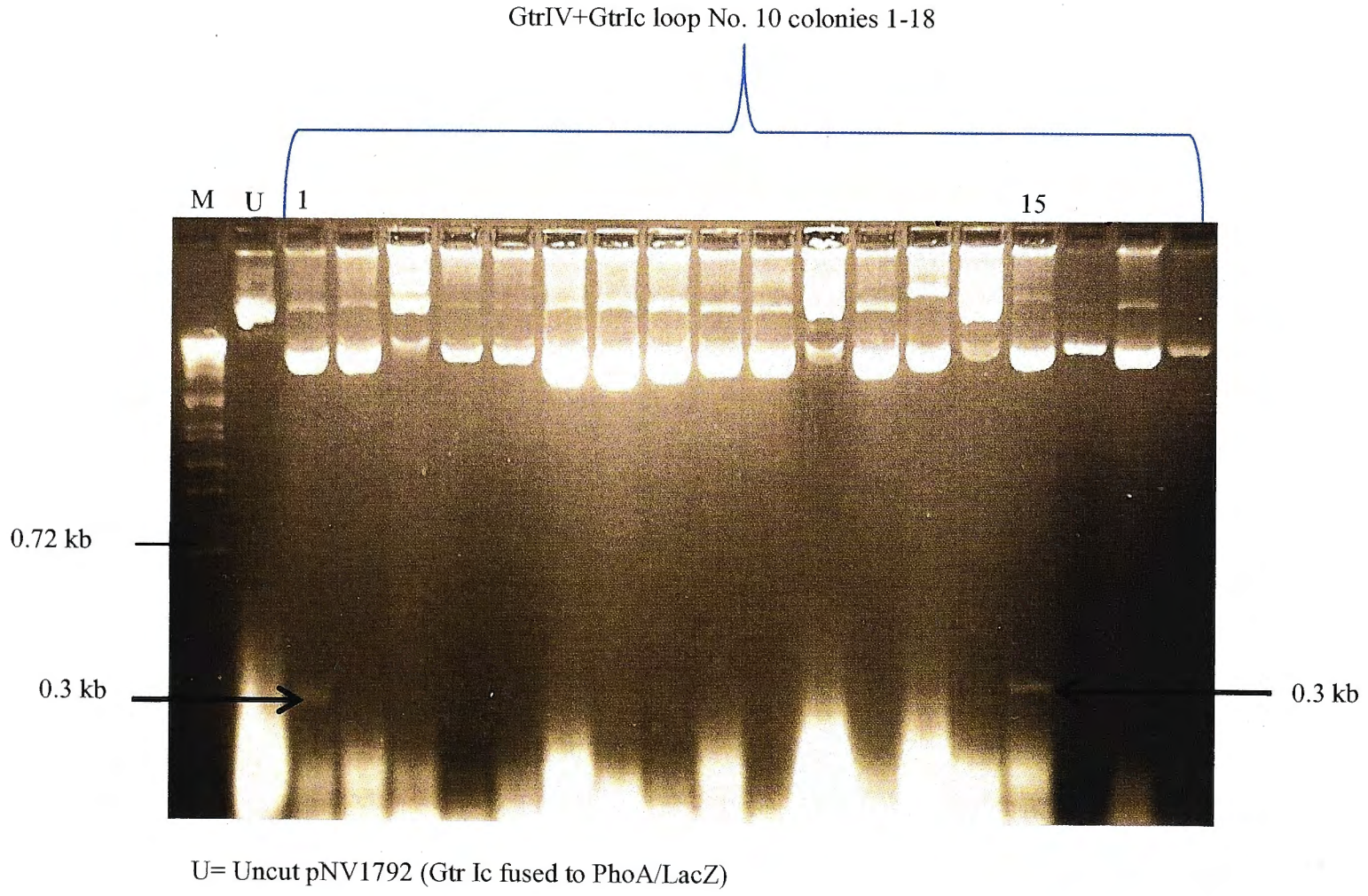


Figure 4.13: Screening of GtrIV-GtrIc loop No. 10-GtrIV chimeras with the help of *Sma*I restriction digests. 18 transformants from the GtrIV-GtrIc loop No. 10-GtrIV transformation plate were digested with *Sma*I. Colonies 1 and 15 displayed two bands at the expected sizes of 6.1 kb and 0.3 kb. The rest of the colonies did not display the smaller 0.3kb band. This meant that they were most probably the vector (GtrIV minus loop No. 6) that self-ligated without the insert (GtrIc loop no. 10). M= Molecular weight marker.

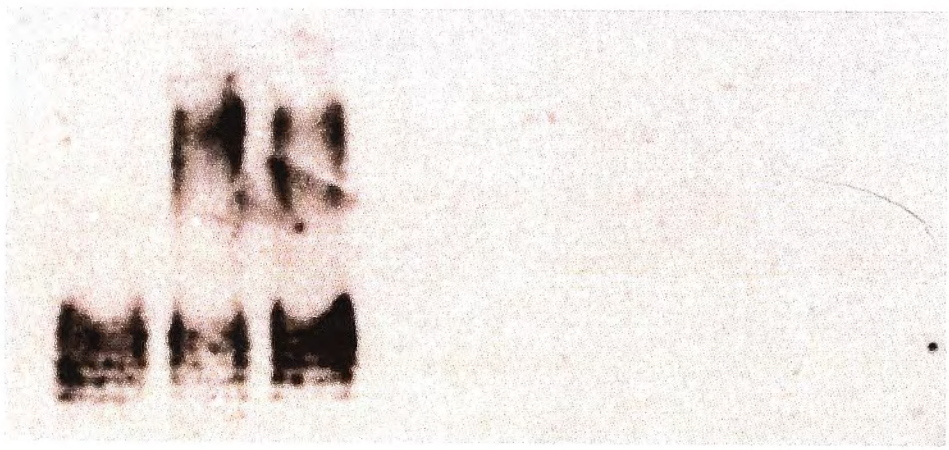
serotype 1c) (Table 4.1). The hybrid GtrIc constructs containing loop No. 2 of GtrIV and loop No. 6 of GtrIV failed to convert serotype 1a to serotype 1c. Similarly, the hybrid GtrIV constructs containing loop No. 2 of GtrIc and loop No. 10 of GtrIc were unable to convert serotype 1a to serotype 1c. To confirm the slide agglutination results, LPS western blots utilising Type IV antisera and MASFIc as the primary anitbodies, were performed (Figure 4.14a and b). Addition of foreign loops to both the Gtrs may have compromised their structural integrity and as a result, prevented them being assembled on the bacterial membrane. This would in turn result in lack of serotype conversion. Therefore, the next step was to determine if these chimeric proteins were assembled in bacterial membrane. To confirm for assembly in the membrane, Western blots (using anti-alkaline phosphatase antibodies) were carried out on the membrane protein extracts of each chimeric protein (Figure 4.14c). Of the four chimeric proteins created, only the GtrIc-GtrIV loop No. 2-GtrIc chimera was not detected by the blot. Its non-functional hybrid GtrIV counterpart however, was shown to have been assembled in the membrane. These results indicated that all the chimera, except for GtrIc-GtrIV loop No. 2-GtrIc were localised in the cell membrane.

4.5 Discussion

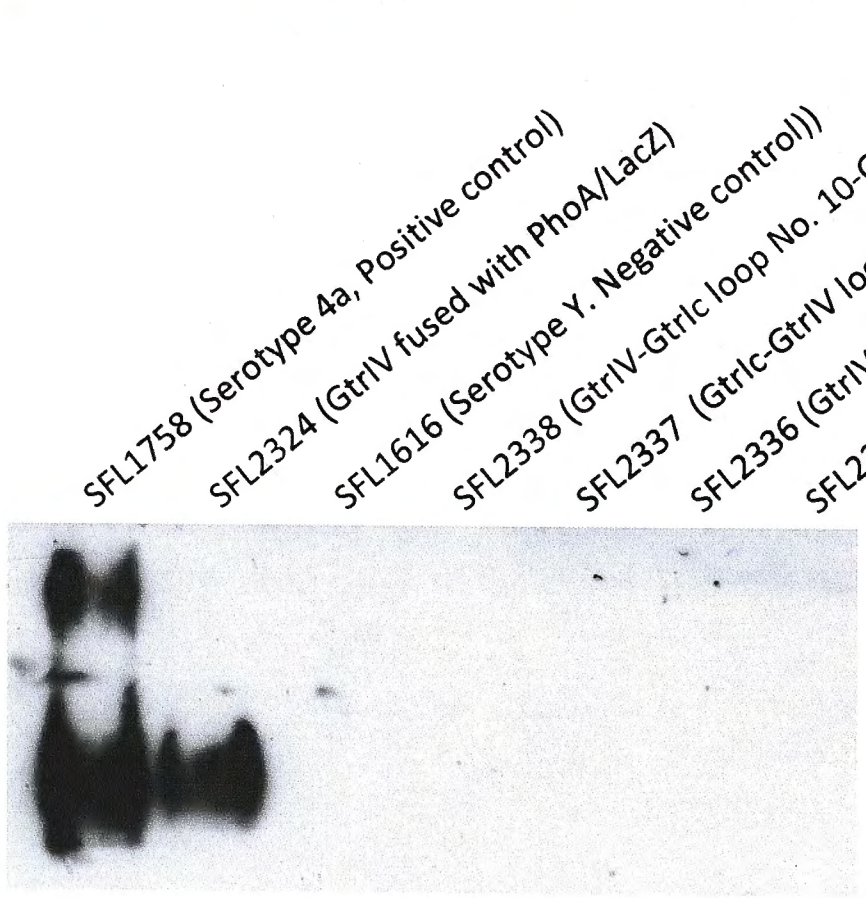
4.5.1 Loop deletions

The recently solved topology of GtrIV has enabled us to further characterize its mode of action by identifying important regions and residues, thus giving us greater insight into how these glucosyltransferases are able to modify the O-antigen and confer serotype specificity. Based on the widely accepted notion that glucosylation is thought to occur in the periplasm, the

a)


SFL1613 (GtrIc Wildtype)
SFL1990 (Serotype Ic (Positive Control)
SFL2069(GtrIc fused with PhoA/LacZ)
SFL1416(Serotype 1a, Negative control)
SFL2339 (GtrIc-GtrIV loop No. 2-GtrIc Chimera)
SFL2340 (GtrIV-GtrIc loop No. 2-GtrIV Chimera)
SFL2341 (GtrIc-GtrIV loop No. 6-GtrIc Chimera)
SFL2342 (GtrIV-GtrIc loop No. 10-GtrIV Chimera)

b)


SFL1758 (Serotype 4a, Positive control)
SFL2324 (GtrIV fused with PhoA/LacZ)
SFL1616 (Serotype Y. Negative control))
SFL2338 (GtrIV-GtrIc loop No. 10-GtrIV Chimera)
SFL2337 (GtrIc-GtrIV loop No. 6-GtrIc Chimera)
SFL2336 (GtrIV-GtrIc loop No. 2-GtrIV Chimera)
SFL2335 (GtrIc-GtrIV loop No. 2-GtrIc Chimera)

c)

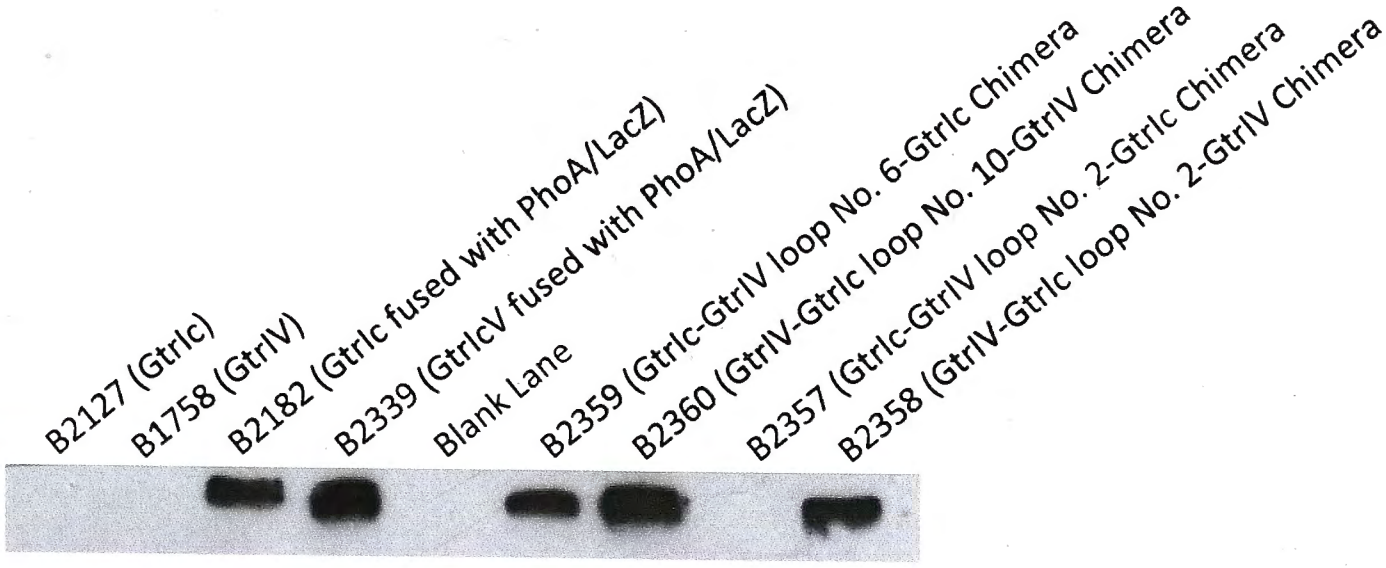

B2127 (GtrIc)
B1758 (GtrIV)
B2182 (GtrIc fused with PhoA/LacZ)
B2339 (GtrIcV fused with PhoA/LacZ)
Blank Lane
B2359 (GtrIc-GtrIV loop No. 6-GtrIc Chimera
B2360 (GtrIV-GtrIc loop No. 10-GtrIV Chimera
B2357 (GtrIc-GtrIV loop No. 2-GtrIc Chimera
B2358 (GtrIV-GtrIc loop No. 2-GtrIV Chimera

Figure 4.14: Western immunoblots of LPS extractions and membrane protein extractions of all chimeras between GtrIc and GtrIV. a) The chimeras were transformed into SFL1416 (serotype 1a) to test for modification of the O-antigen to serotype Ic by probing with MASF Ic. All the chimeras were non-functional. b) Chimeras between GtrIC and GtrIV were also transformed into SFL1616 (serotype Y) to test for modification of O-antigen to serotype 4a by probing with Type IV antisera. All chimeras were found to be non-functional. c) Western immunoblots of membrane protein extracts of all chimeric constructs. GtrIc fused to PhoA/LacZ is 111 kDa while GtrIV fused to PhoA/LacZ is 108kDa. With the exception of B2357, the membrane extracts of all the other hybrid constructs show bands similar to the positive controls, thus indicating localisation of the chimeras in the bacterial membrane.

two large periplasmic loops of GtrIV (loop No. 2 and loop No. 6) were targeted to investigate their role in GtrIV function. At the same time, GtrIV was found to have a large 30 amino acid spanning cytoplasmic loop (loop No. 3). As such a loop is not found in other members of the $Gtr_{(type)}$ family of *Shigella flexneri*. It was also targeted for functional studies.

A total of 6 loop deletions were performed on GtrIV to help identify critical regions within the protein. The loop No. 2 and loop No. 3 deletion proteins were found to have abolished GtrIV function. Membrane protein western blots subsequently revealed that both deletion proteins were not present in the cellular membrane. This effectively meant that the deletions prevented GtrIV from being assembled in the cellular membrane. The deleted loops could have been important for the structural integrity of the GtrIV and their deletion could have led to the disruption of van der Waals forces thus affecting the hydrophobicity of the mutant protein (Fiedler *et al.*, 2010). This would have resulted in protein misfolding, thereby potentially causing hydrophobic regions within the protein to be exposed to the extracellular environment. Consequently, such an occurrence would likely prevent the protein from being inserted into the membrane. Harrington and Ben-Tal (2009) have stated that the stability of the tertiary protein structure is highly dependent on inter-helical interactions. These include hydrogen bonds, salt bridges, aromatic interactions and closely packed residue interaction help stabilise the tertiary protein structure. Therefore, it can be assumed that removing loop No. 2 and No. 3 could have disrupted these inter-helical interactions and led to an unstable, misfolded protein. These aberrant proteins would then aggregate within the cytosol and be targeted for degradation by cellular proteases (Martínez-Alonso *et al.*, 2009). The deleted segments could also have played a role in helping GtrIV localise in the membrane. It is known that transmembrane proteins contain signal

sequences that function as both the targeting and recognition signal for the cell's translocation machinery (Driessen *et al.*, 2001, Fekkes & Driessen, 1999). Found in the N-terminal and C-terminal regions of the protein, the signal sequences have an impact on the overall charge of the N-terminal and C-terminal regions of the protein. This overall charge is then recognised by cellular transport machinery which translocates the protein to its designated location in the cell (Fekkes & Driessen, 1999). Hence, by deleting loops No. 2 and No. 3 from GtrIV, the overall charge of the N-terminal region may have been modified preventing localisation of the mutant protein to the membrane.

Though all the loop No. 6 deletions also prevented GtrIV from functioning, the weak signals obtained from their membrane protein western blots (loop No. 6 deletion and two of the three loop No. 6 partial deletions) indicated that there was a small amount of protein localised on the bacterial membrane as compared to the fully intact protein. Again, it can be seen that the deletions would have caused a conformational change in GtrIV, thus affecting its native ability to localise in the membrane. This would in turn lead to a negligible amount of O-antigen modification, too little to be picked up in the LPS western blots. Also, such a conformational change could have rendered the proteins to be non-functional anyway. However, in B2351 (loop No. 6 partial deletion 3) the amount of protein that was assembled in the bacterial membrane was almost as high as the amount of the intact GtrIV. This, coupled with the loss of functionality of the deleted protein, provides us with an insight that the missing 19 amino acid segment of loop No.6 partial deletion 3 could contain residues that may be part of a catalytic site present in the large periplasmic loop No. 6. This led us to carry out the 4 further deletions to zero in on the exact region within these 19 amino acids that could potentially be the catalytic site. Slide

agglutination and LPS westerns revealed that FD 1 and 2 yielded functional proteins while FD 3 and 4 resulted in non-functional proteins. On the other hand, western blots of the membrane proteins extracted from these constructs show that all the four further deletion proteins were localised in the membrane. This is in agreement with the hypothesis that the missing segments may be part of a catalytic site present in the large periplasmic loop No. 6 that might not only be responsible for interaction with the glucosyl residue to be attached to the O-antigen, but also with potential interactions between GtrIV, GtrA and GtrB. As such, the next step would be to carry out site directed mutagenesis experiments that target all the amino acids that are within FD 3 and 4. This will allow us to pinpoint critical residues within this segment. Identification of such residues will help us in understanding the mechanism of action of GtrIV.

4.5.2 Chimeric proteins between GtrIc and GtrIV

Chimeric proteins were created between GtrIV and its closet structural homologue, GtrIc. They were created in an attempt to test the well held hypothesis that topologically similar N-terminal periplasmic loops in all $Gtr_{(type)}$ proteins may perform the conserved functions of interacting with the UndP-glucose and possibly recycling it to the cytoplasmic face. In light of this, two loop No. 2 chimeras were created by swapping loop No. 2 of GtrIV with that of GtrIc and vice-versa. However, slide agglutination and LPS western blots have revealed that both chimeras had lost their native function. Membrane protein western blots showed that GtrIc-GtrIV loop No. 2-GtrIc chimera was not detected by the blot, unlike its GtrIV counterpart. This indicates that although loop No. 2 is topologically conserved, its low sequence identity cannot necessarily establish a functional protein as observed in the GtrV-GtrX Loop No .2-GtrV chimera (Korres and Verma, 2006). In retrospect, it is possible that the loop No. 2 of both

proteins may contain unique amino acid residues that are critical for its conserved interaction with the UndP-Glucose precursor, owing to the fact that the active sites for both these proteins might require slightly different tertiary structures in order to interact specifically with the O-antigen (Nair *et al*, 2011). For the GtrIV-GtrIc loop No. 2-GtrIV chimera, the results imply that specific interactions between residues present in its native loop No. 2 (of GtrIV) are required to form a catalytic site that facilitates O-antigen modification. On the other hand, the absence of assembly in the bacterial membrane of the GtrIc-GtrIV loop No. 2-GtrIc chimera suggests that GtrIV loop No. 2 has compromised the structural integrity of the GtrIc protein which would have resulted in a misfolded protein that was unable to be translocated to the plasma membrane and subsequently, would have most probably been degraded.

The large C-terminal periplasmic loops of GtrIV and GtrIc (loop No. 6 and loop No. 10 respectively) were also swapped between the two proteins to test the hypothesis that these loops are responsible for conferring serotype specificity of their respective Gtrs. Slide agglutination and LPS Western blots revealed absence of modification to respective serotypes. Membrane protein western blots revealed that both hybrid proteins were localised in the cellular membrane. As both large periplasmic loops contain over 100 amino acids each (123 amino acids in loop No. 6 of GtrIV and 120 amino acids in loop No. 10 of GtrIc), each loop can possibly accommodate its own unique tertiary structure that would contribute to the serotype specificity of O-antigen modification. Therefore, in absence of this, GtrIV and GtrIc would not be able to carry out their specific functions. Alternatively, the structural integrity of each chimeric protein may have been compromised with the addition of the foreign loop. The loss of function for GtrIc-GtrIV loop No. 6-GtrIc and GtrIV-GtrIc loop No.10-GtrIV chimeras can also be attributed to the importance of

the two C-terminal periplasmic loops in conferring serotype specificity through the addition of glucosyl residues to the O-antigen in a site and linkage specific manner. As both loops span more than 100 amino acids each, this would facilitate intra-loop and inter-loop interactions that may contribute to serotype specificity and O-antigen modification.

4.6 Conclusion

Based on the newly elucidated topology of GtrIV, the roles of the N-terminal and C-terminal periplasmic loops of GtrIV (loop No. 2 and loop No. 6) were investigated by carrying out a series of loop deletions. As GtrIV was also found to contain an unusual (compared to the other Gtrs) cytoplasmic loop No. 3, it was also deleted to investigate its function. It was found that the loop No. 2 and No. 3 deletions knocked out GtrIV function by preventing the protein from assembling in the bacterial membrane. By sequentially deleting loop segments in loop No. 6, the presence of a potential catalytic site located between resides D260 to W269 was hypothesised. To further investigate the roles of conserved or specific functions of the two periplasmic loops of GtrIV, loop swap experiments between the N-terminal periplasmic loops and the C-terminal periplasmic loops were undertaken. The resulting hybrids lost their native function and were unable to substitute function to the other protein. This signifies the importance of both loops in GtrIV function.

Chapter 5

Identification of residues that are critical for GtrIV function

5.1 Introduction

As mentioned in chapter 4, functional analysis of a protein helps to provide an insight into areas responsible for its mechanism of action. As such, residues critical for the function of a protein can be identified from data gained as a result of mutations that lead to a change of the protein's function (loss of function or hyper-function). The ability to change a specific amino acid and thereby modulate catalysis has been invaluable in determining which residues are directly involved in a reaction. Site-directed mutagenesis employed in functional studies involves amino acid replacement, deletion and insertion of varying lengths of amino acid stretches (Alexeyev & Winkler, 2002). It has allowed the consequences from a wide array of side chain substitutions to be assessed and has been instrumental, in conjunction with other techniques, in unraveling energetic, functional, structural, and dynamic properties of the protein matrix (Brockhausen *et al.*, 2008). Only a small fraction of the residues of an enzyme directly participate in catalysis. These residues that are critical may not only be involved directly in catalysis but also in interacting with substrates and other proteins in a complex (Brockhausen *et al.*, 2008). They may also be responsible for maintaining the structural integrity of the protein such that other critical residues may form the catalytic site (Alexeyev & Winkler, 2002).

The identification of critical residues in GtrIV is being reported for the first time. Previous studies have highlighted critical residues in GtrII and GtrV. Lehane *et al.,* (2005) identified 4 residues (E40, F414, C435 and K478) that abolished function in GtrII when mutated to alanine. Furthermore, a previous study by Chen *et al.*, (2003) identified residue C437 to be

critical in GtrII. As for GtrV, Korres & Verma (2006) reported that two N-terminal residues (E42 and D43) and one C-terminal residue (D380) were found to be critical.

5.2 Previous work

Despite the fact that a number of residues have been targeted, thus far, site-directed mutagenesis work on GtrIV has not yet uncovered any critical residues. The aspartic acid residues D23, D31 and D33 located in the periplasmic loop No. 2 of GtrIV were targeted based on a series of alignments showing that these residues are conserved among all the Gtrs. D23 was found to be conserved between in all Gtrs as shown by Lehane *et al,* (2005) (Figure 5.1a). Residue D31 was chosen on the basis that it may work together with D33 as a kind of a motif as seen in other glucosyltransferases which form the catalytic site (Breton & Imberty, 1999) (Figure 5.1b). Furthermore, some alignments showed that D31 lines up with critical residues found in other Gtrs. D33 was found to be conserved when another alignment was carried out by Korres & Verma (2006) (Figure 5.1c). Residues D31 and D33 were then mutated in the same construct to investigate whether they worked in tandem with each other. The resulting mutant tested positive for serotype conversion indicating that GtrIV was still able to function (Nair A, Honours thesis 2006).

Several other residues located within GtrIV loop No. 6 were also targeted as it was thought that this loop may be responsible for modifying the O-antigen in a sugar and linkage specific manner. Single point mutations were introduced on F273, D276 and C288. F273 and D276 were targeted solely on the fact that they lined up with a critical phenylalanine F414 in GtrII and an aspartic acid residue D380 in GtrV respectively (Figure 5.2a and b). Both mutations

(a)

```
GtrI      MSICIKQSAL-KI--LLAL---S--ALLIT--W--LTTRYFPVEP-DV
GtrII     MI---KINNLFKNANLLA--FISCFAISIYCYWGWLYDGTLNIDG-EF
GtrIV     MAE-R---I-RYVF--LAI-FCAL-LC---K-----------ESIDI
GtrV      MKSLKT-SYVKKLFF-LIGVFFIAVILMYLRRPDIIRAPQFWAE--DA
GtrX      MRNWHKISIFI-LAFTLIW-'---------LRRIDILTNAQFWAE--DA
```

D23

(b)

```
                1          11          21          31          41          51
GtrV      IAVILMYLRRPD-IIRAPQFW---AED----AHVWYAMAYNNGIFTSM---IFPQNGYYQ
GtrX      ---------RID-ILTNAQFW---AED----AVFWYKDAYEQGFFSSL---ITPRNGYFQ
GtrI      ---TRYFPVEPD-VANSPIVWRHILENGISSIHDWKP-TVDNWYFT-----VYPINFLFY
GtrII     ISIYCYWGWLYDGTLNIDGEF---TNN------FYQTITLGRWFHTFLRHYFLPEPFSLY
GtrIV     -SIDIWMMNNGD-FDRAITPF---MNG------IRQFSHDGTLLYTLKENFSSIVNYEYK
```

D31

```
          61        71
GtrV      TISKLIASISL---------
GtrX      TVSTLIVGATTFINPIYAPL
GtrI      ML--LGDDG-----------
GtrII     ITP-LIALSF----------
GtrIV     SSFSYILYLYA---------
```

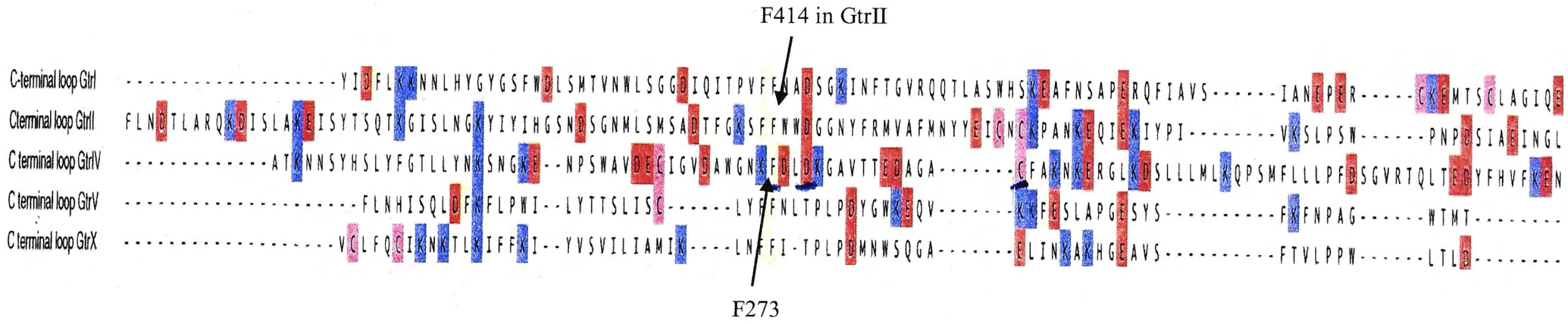
(c)

```
GtrI      YFPVEPDVANSPIVWRHI
GtrII     DGTLNIDGEFTNNFYQTI
GtrIV     MNNGDFDRAITPFMNGIR
GtrV      PQFWAEDAHVWYAMAYNN
GtrX      AQFWAEDAVFWYKDAYEQ
```

D33

Figure 5.1: (a) Manual alignment of the N-terminal sequences (loop No. 2) of *S. flexneri* Gtr proteins (Lehane *et al.,* 2005). The positions of the conserved acidic residues are shown in red. (b) An alternative alignment of the N-terminal sequences of all the Gtrs shows that D31 in GtrIV is conserved. Acidic residues are in red. (c) A best fit analysis performed by Korres and Verma (2006) of the first periplasmic loop No. 2 present in all Gtrs. The conserved aspartic acid (D) is shown in red. Acidic residues near the conserved residues are indicated in bold.

(a)


F414 in GtrII
C-terminal loop GtrI
Cterminal loop GtrII
C terminal loop GtrIV
C terminal loop GtrV
C terminal loop GtrX
F273

(b)

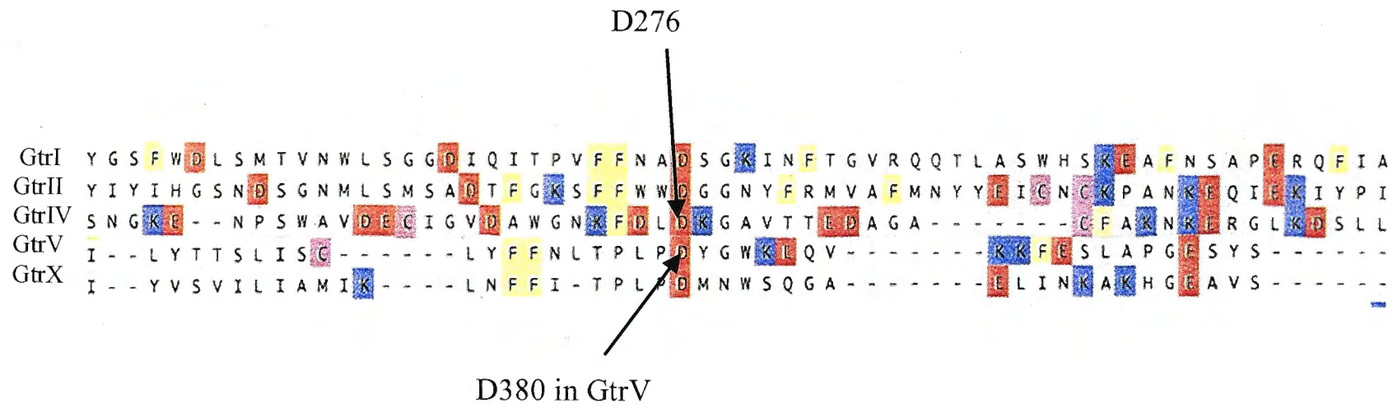

D276
GtrI
GtrII
GtrIV
GtrV
GtrX
D380 in GtrV

Figure 5.2: C-terminal alignments of the various Gtrs. (a) F273 of GtrIV lines up with F414 of GtrII (phenylalanine residue found critical in GtrII). (b) A manual alignment shows that D276 of GtrIV lines up with D380 of GtrV (critical aspartic acid residue).

did not affect GtrIV function (Nair A, Honours thesis 2006). Previously, Lehane *et al.,* (2005) and Chen *et al.,* (2003) had shown that cysteine residues C435 and C437 in the large periplasmic C-terminus of GtrII, are essential for function. Thus, C288 was selected as it was thought to form a disulphide bond with C263 in GtrIV loop No. 6 and was non-critical.

Apart from mutating single residues, to further investigate the importance of the negatively charged acidic residues occurring within loop No. 6, three sets of paired residues DE261, ED283 and ED326 were mutated to double alanines (AA261, AA283 and AA326) by site-directed mutagenesis. Functional analysis of all constructed mutants showed agglutination to Type IV antiserum when introduced into SFL1616. A full summary of the residues already targeted, along with the reasoning behind their selection is shown in Table 5.1. Figure 5.3 shows the location of the residues in GtrIV. All these residues are non-critical for GtrIV function (Nair A, Honours thesis 2006)

5.3 Acidic residues in GtrIV

It is postulated that acidic residues might interact with UndP-Glucose (N-terminal periplasmic loop) or involved in attaching the glucosyl group to the specific rhamnose of the O-antigen (C-terminal periplasmic loop) (Korres & Verma, 2006). As there are a total of 20 acidic residues located within the periplasmic loops No. 2 (5 residues) and No. 6 (15 residues) of GtrIV, the first step was to target all the remaining acidic residues within the two large periplasmic loops individually (Figure 5.4) In loop No. 2, the three remaining residues D49, E58 and E67 were targeted. Likewise, all the individual acidic residues in loop No. 6 were targeted (E254, D267, D274, E294, D299, D317, E334 and E347). The template DNA, pNV1917

Table 5.1: Residues targeted by site-directed mutagenesis based on bioinformatic analysis. The rationale for choosing each residue and the primers used for each mutation are also listed below.

Residue	Rationale	Primers
D23	An alignment done by Lehane *et al.,* (2005) showed that it lined up with a critical aspartic acid residue on loop No.2 of GtrI.	D23ForA D23RevA
D31	Negatively charged residue found close to D33. Found to line up with critical aspartic acid residues of other Gtrs in some alignments. May work in tandem with D33 for function.	D31ForA D31RevA
D33	A conserved residue present in loop No 2 of all the Gtrs	D33ForwardA D33ReverseA
F273	Lines up with a critical F residue found in the C-terminus of GtrII	F273ForwardA F273ReverseA
D276	Lines up with a critical D residue found in the C-terminus of GtrV (H. Korres, PhD Thesis, 2006)	D276ForwardA D276ReverseA
C288	Thought to form a disulphide bond with C263.	C288ForwardA C288ReverseA
DE261	D and E paired residues found in Loop No. 6	DE261ForAA DE261RevAA
ED283	E and D paired residues found in Loop No. 6	ED283ForAA ED283RevAA
ED326	E and D paired residues found in Loop No. 6	ED326ForAA ED326ForAA

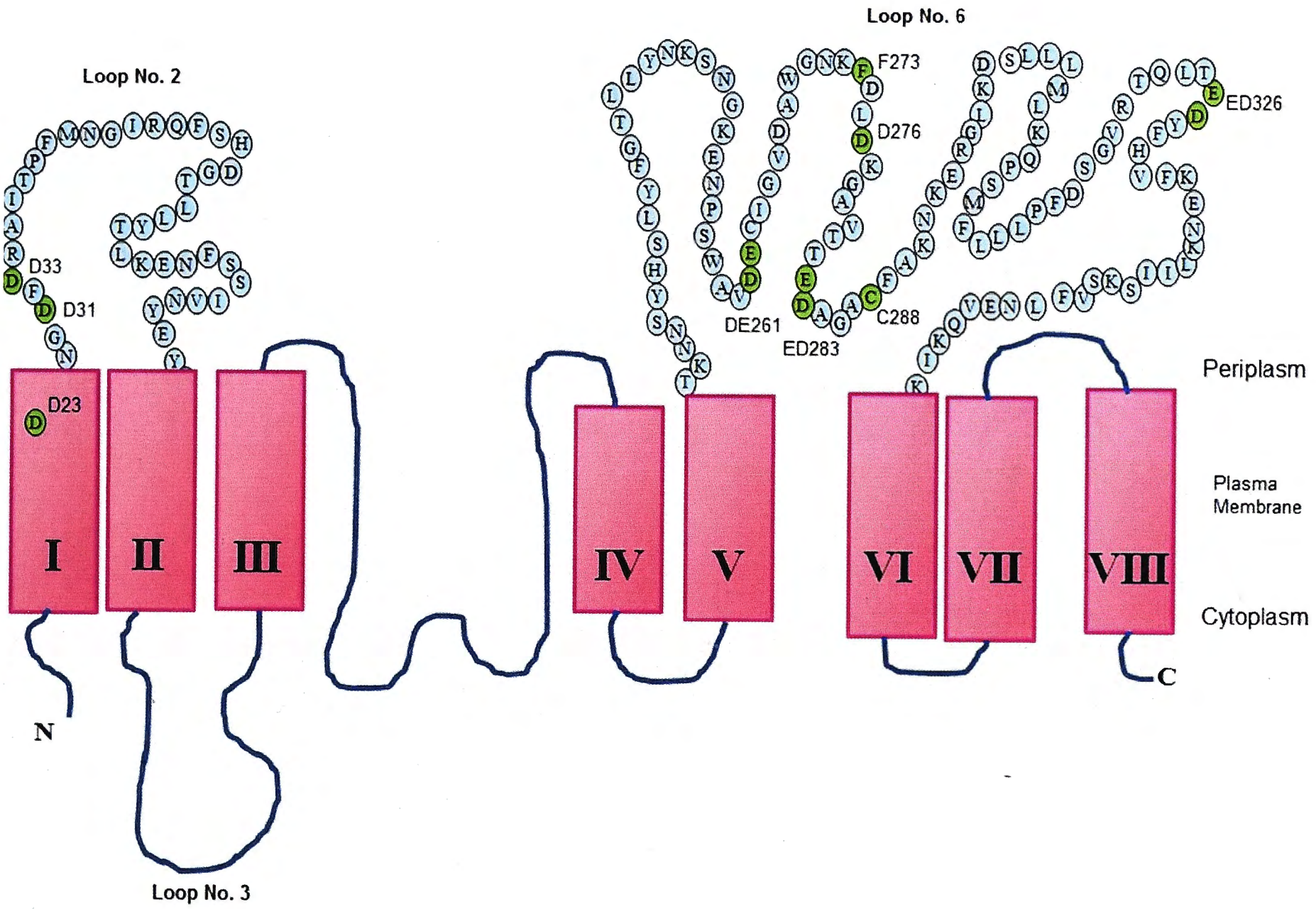


Figure 5.3: Model of GtrIV showing only periplasmic located residues. Highlighted residues were targeted by site-directed mutagenesis based on multiple alignments of all Gtrs. Residues D31 and D33 are in loop No. 2 while F273, D276 and C288 are located within loop No. 6.

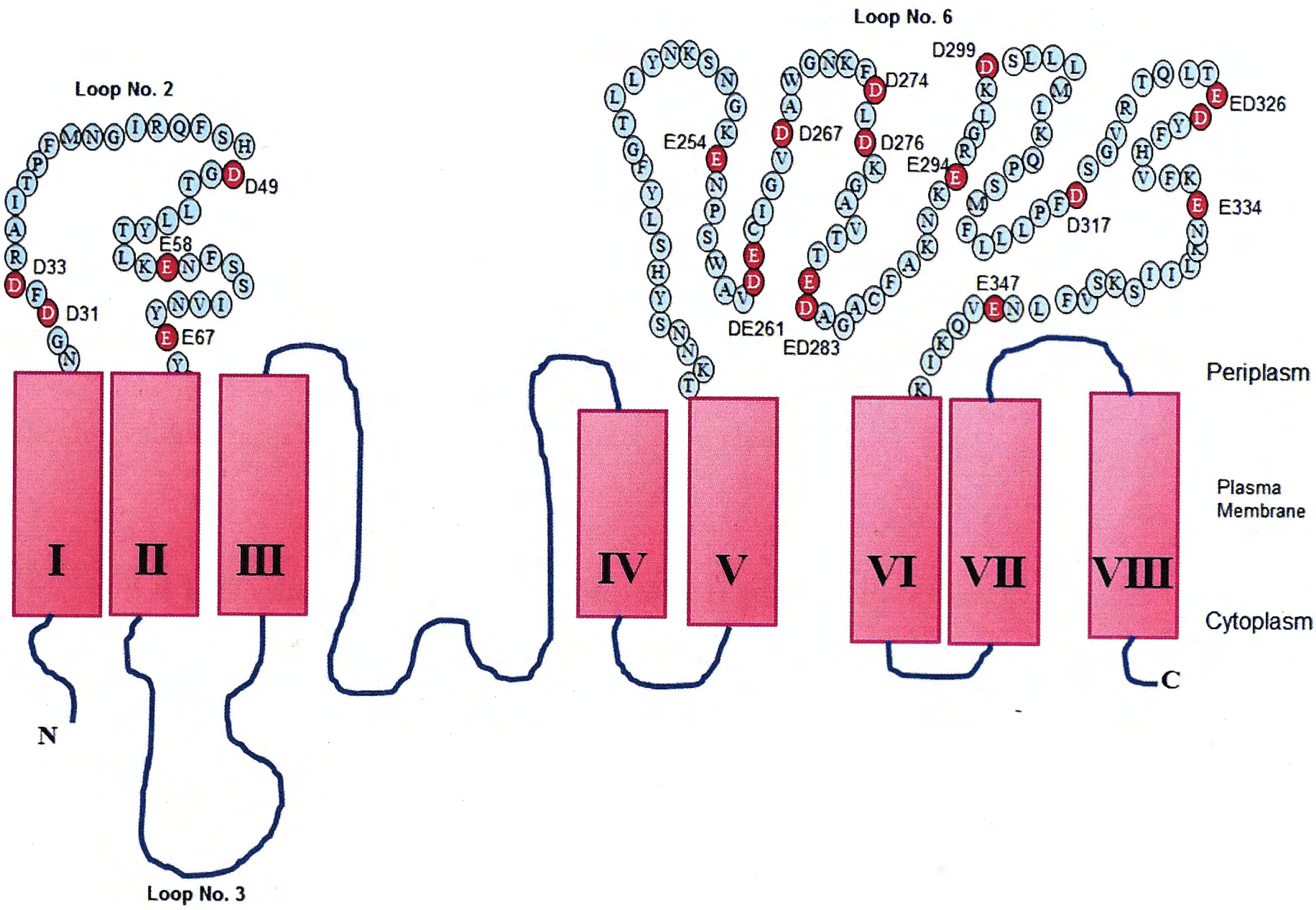


Figure 5.4: Negatively charged residues found on loop No. 2 and loop No. 6 (in pink). A total of 20 negatively charged residues are present. In loop No. 2, residues D49, E58, E67 were not targeted previously and therefore, were targeted in this study. The number of negatively charged residues scattered throughout loop No. 6 seems to suggest that these residues may have a compensatory effect in the event of a mutation involving negatively charged residues. Negatively charged residues that occur in pairs are DE261, ED283 and ED326.

carrying *gtrIV* fused to *phoA/lacZ* was used for all site-directed mutagenesis reactions. Each residue was mutated to an uncharged alanine. Upon confirmation of DNA amplification, the DNA was treated with *Dpn*I, transformed into electrocompetent XL1-Blue cells. DNA from 3 resulting colonies taken from each transformation plate was extracted and sent for sequencing. Constructs encoding the mutated proteins were then transformed into SFL1616 and transformants were tested for agglutination with *S. flexneri* Type IV antiserum. All mutants were able to convert serotype Y to serotype 4a, indicative of a functional GtrIV. Table 5.2 lists all the acidic residues targeted and their slide agglutination results.

5.4 Residues within GtrIV loop No. 6 partial deletion 3

The difference in structure between GtrIV and the other Gtrs (except GtrIc) could mean that other residues may be involved in forming the catalytic site either independently or in tandem with the residues that were initially selected. The results of the loop deletion experiments carried out in Chapter 4 has helped narrow down the region of the protein which is directly involved with its catalytic activity. It has highlighted a distinct possibility that this 19 amino acid spanning region may contain 1 or more critical residues. Therefore, instead of resorting to random selection of other residues within GtrIV loop No. 6, this region was now targeted to identify the amino acids of interest residing within it (Figure 5.5).

5.4.1 Site-directed mutagenesis of all individual residues within further deletions 3 and 4

GtrIV further deletions 3 and 4 (FD 3 and 4) (see Chapter 4) were shown to have knocked out the serotype converting ability of GtrIV (Nair *et al.,* 2011). Thus, all the amino acids that made up these two further deletions were mutated to alanine individually and their

Table 5.2: List of the remaining acidic residues within GtrIV that were targeted and their slide agglutination results. All mutants exhibited agglutination against *S. flexneri* Type IV antisera.

Construct	Agglutination (Type IV)
Loop No. 2 Residues	
D49	+
E58	+
E67	+
Loop No. 6 Residues	
E254	+
D267	+
D274	+
E294	+
D299	+
D317	+
E334	+
E347	+

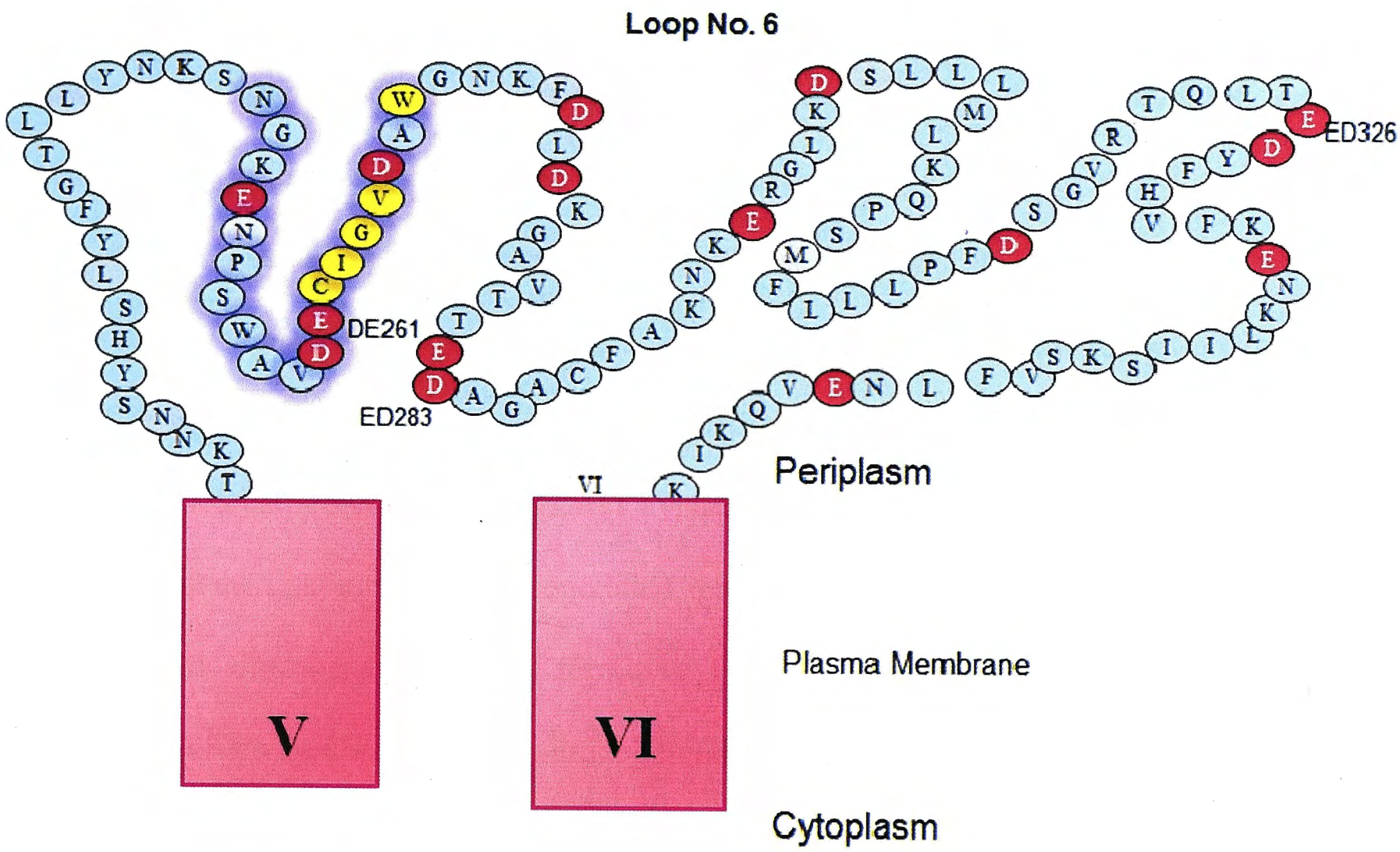


Figure 5.5: Residues located within GtrIV loop No. 6 partial deletion 3. The 19 residues within the purple highlights make up GtrIV loop No. 6 partial deletion 3. Residues in yellow were the individual residues that were targeted for site-directed mutagenesis. They are C263, I264, G265, V266 and W269. A268 was not targeted as it is already an alanine. All acidic residues are shown in pink.

effects on GtrIV were studied. The amino acid targets are shown in Figure 5.5. Of the 9 amino acids within FD 3 and 4, only C263, I264, G265, V266 and W269 were mutated. A268 was not mutated as it is an alanine while DE261 and D267 have already been mutated. All amino acids were mutated to alanines via the same site-directed mutagenesis protocol described above. The template used was pNV1917. Once the mutations had been confirmed, functional analysis was carried out by transforming the mutated constructs into SFL1616. Following this, slide agglutination was carried out with *S. flexneri* Type IV antiserum and agglutination was observed for all the mutants. As such, it was concluded that the amino acids within this region, when mutated individually, were not critical for GtrIV function. A summary of all the residues that were mutated within GtrIV loop No. 6 partial deletion 3 is shown in Table 5.3.

5.4.2 Cumulative mutation of acidic residues within GtrIV loop No. 6 partial deletion 3

Because of the large number of acidic residues present within GtrIV loop No. 6. There is a chance that they work in tandem with one another. Also, the other acidic residues could act to compensate for the loss of any other acidic residue within this large periplasmic loop. To investigate this, the acidic amino acids within GtrIV loop No. 6 partial deletion 3 were mutated in combination of one another (Figure 5.5). There were a total of 4 acidic residues in this region, namely, E254, D261, E262 and D267. Seeing that the paired residue mutant DE261 (pNV1958) was still functional, it was used as a template to knock out E254 and D267 in tandem. Confirmation of the mutants resulted in two new mutant constructs E254 + DE261 (pNV1964) and DE261 + D267 (pNV1965). In parallel, E254 was mutated in tandem with D267. The template used for this site-directed mutagenesis PCR was pNV1950 (D267 mutant). Sequence confirmation of resulting clones led to the creation of pNV1966 (E254 + D267). All three

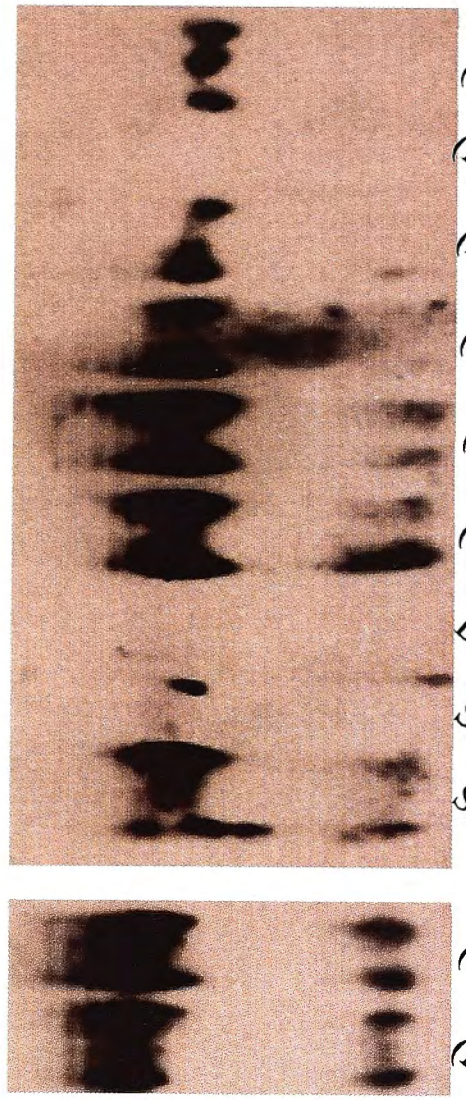
Table 5.3: Table of residues located within GtrIV loop No. 6 partial deletion 3 that were mutated to alanine (neutral charge).

Construct	*Agglutination (Type IV)*	*Agglutination (Group 3, 4)*
Single Mutants		
E254	+	+
C263	+	+
I264	+	+
G265	+	+
V266	+	+
D267	+	+
W269	+	+
Double Mutants		
DE261	+	+
E254+D267	+	+
D261+D267	+	+
E262+D267	+	+
Triple Mutants		
E254+DE261	+	+
DE261+D267	-	+

constructs were then transformed into SFL1616 and tested for agglutination with *S. flexneri* Type IV antiserum (Table 5.3). Agglutination was observed for the mutants E254 + D267 (SFL2291) and E254 + DE261 (SFL2294). However, no agglutination was seen for the mutant DE261 + D267 (SFL2305). As mentioned in chapter 4, there is a possibility that decreased protein function makes agglutination hard to visualise (though it may be present). Therefore, bacterial LPS was extracted from these mutants and western blots using Type IV antisera was performed on them (Figure 5.6a). The western blots confirmed that the DE261 + D267 mutant was non-functional. This result meant that the function of GtrIV was affected when 3 acidic residues within the targeted region were knocked out in tandem.

The possibility that only two out of the three residues may be sufficient for knocking out GtrIV function was explored. As DE261 was mutated as a pair, the effect of each residue mutated individually (D261 and E262) in tandem with D267 was examined. To obtain these mutants, site-directed mutagenesis was carried out using pNV1965 (D267 + DE261 triple mutant) as a template. In this case, specific mutagenic primers were used to restore the mutated residues back to the original amino acid to obtain the double mutants. Therefore, to obtain a D261 + D267 double mutant, mutagenic primers GtrIVA262EFor and GtrIVA262ERev were designed such that A262 in the triple mutant was reverted back to E262, thus rendering the construct with mutations at D261 and D267. In the same vein, the E262 + D267 double mutant was obtained with help of the mutagenic primers GtrIVA261DFor and GtrIVA261DRev that reverted A261 back to D261. Confirmation of the mutants resulted in two new mutant constructs D267 + D261 (pNV1988) and D267 + E262 (pNV1990). Subsequent functional analyses of these constructs were carried out by transforming them into SFL1616 and testing for

a)



(E254+DE261)
(DE261+D267)
(E254+D267)
(D267)
(DE261)
(E254)
Blank Lane
SFL1616 (Serotype Y, -ve control)
SFL2324 (GtrIV + PhoA/LacZ, +ve control)

(E262+D267)
(D261+D267)

b)

(GtrIV + PhoA/LacZ, +ve control)
(GtrIV, -ve control)
Blank Lane
(E254)
(DE261)
(D267)
(E254+DE261)
(DE261+D267)
(E254+D267)

(D261+D267)
(E262+D267)

Figure 5.6: Functional analysis of all mutations involving aspartic acid and glutamic acid residues that occur in GtrIV loop No. 6 partial deletion 3 (N251 to W269). a) LPS Western blot results. All mutant constructs were transformed into a serotype Y strain (SFL1616) to test for modification of the O-antigen to serotype 4a. SFL2324 is GtrIV fused to PhoA/LacZ and then transformed into SFL1616. The binding of LPS to the antibody is indicative of serotype conversion from serotype Y to serotype 4a. All mutants, except for the DE261+D267 triple mutant, were positive for serotype conversion to serotype 4a when probed with Type IV antisera. The LPS western blots confirm that the DE261+D267 triple mutations abolish GtrIV function and this cease in function only occurs when all three residues (D261, E242 and D267) are mutated in tandem. b) Western blots (using anti-alkaline phosphatase primary antibody) performed on the membrane protein extractions of the various GtrIV point mutants that were fused to alkaline phosphatase. All the mutant constructs were detected when the membrane extracts were probed with anti-alkaline phosphatase. This indicated that the mutant proteins (which are fused to PhoA/LacZ) are expressed and localized in the cellular membrane.

agglutination with *S. flexneri* Type IV antiserum (Table 5.3). Agglutination was observed for both mutants. This result was then confirmed by LPS western blots (Figure 5.6a)

To determine whether the D267 + DE261 triple mutant had an effect on protein assembly, western blots using anti-alkaline phosphatase antibodies were carried out on the membrane protein extracts on all the double and triple mutants (Figure 5.6b). The membrane protein western blots revealed that, as expected, all the mutants were found to be assembled on the bacterial membrane. This result indicates that the DE261 + D267 triple mutant had abolished the function of GtrIV without compromising its ability to localise in the bacterial membrane.

5.5 Investigation into critical residues within the re-entrant loop

Previous studies have predicted that the re-entrant loops present in members of the Gtr family are important in structural stability of their respective Gtrs. Furthermore, it is believed that the re-entrant loops confer flexibility to the Gtr proteins allowing their respective periplasmic N-terminal and C-terminal regions to come into contact and interact during serotype conversion (Korres & Verma, 2004, Korres & Verma, 2006). It was recently discovered that a proline residue, P194 of GtrIc, when mutated, led to a non-functional GtrIc (Wang A, Honours thesis 2011). P194 is the only proline residue present within the GtrIc re-entrant loop. As GtrIV is the closest structural homologue to GtrIc, it was found that a proline residue P154 was present within the re-entrant loop of GtrIV (Figure 5.7). As such, this residue was mutated to investigate if it plays the same role in GtrIV as it does for GtrIc.

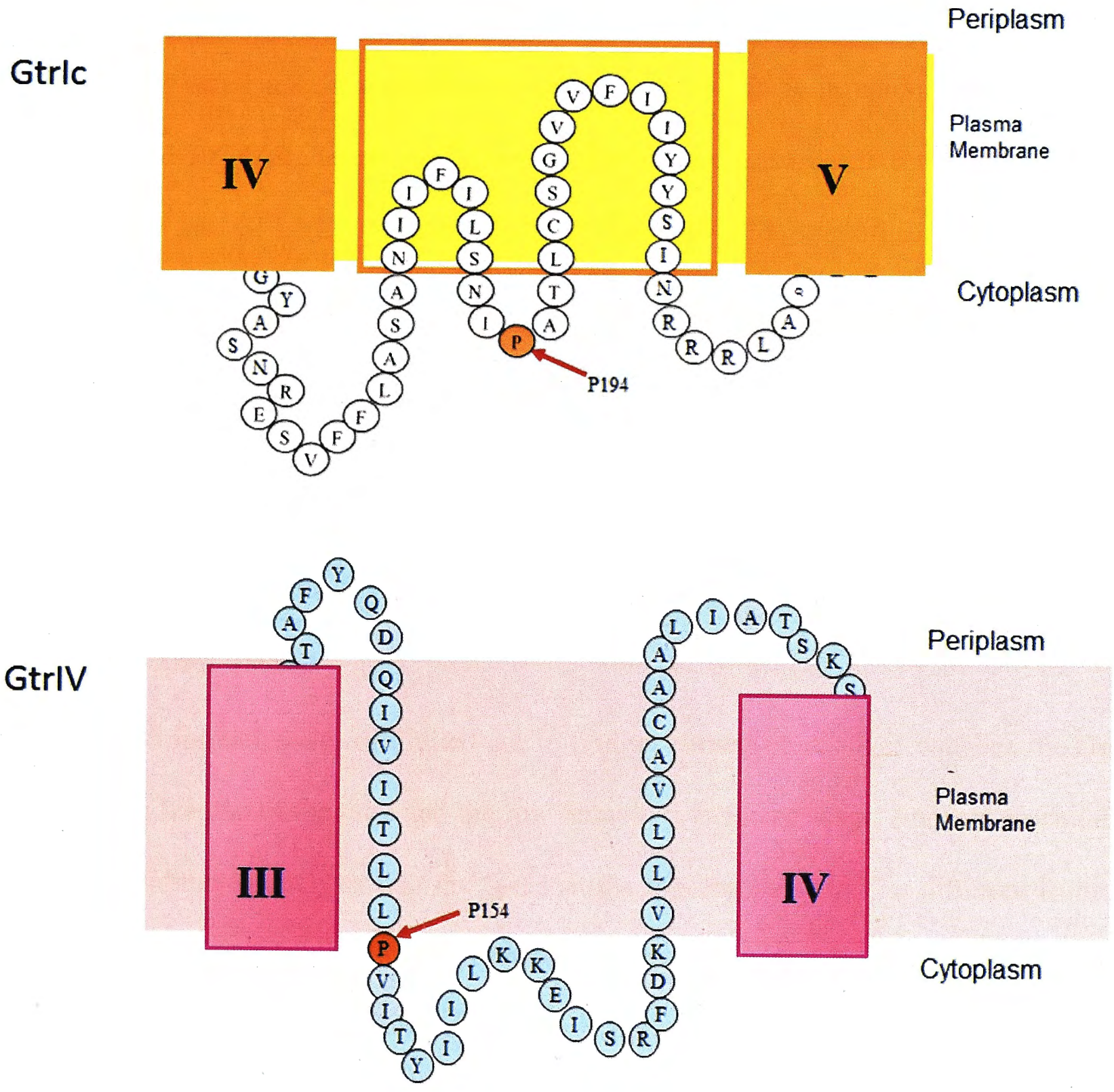


Figure 5.7: Residues located within the re-entrant loops of GtrIc and GtrIV. P194 is the only proline residue present in the re-entrant loop of GtrIc. Likewise, it can be seen that P154 is the only proline residue present within the GtrIV re-entrant loop.

By carrying out the same site-ditected mutagenesis protocol as above, pNV1917 (GtrIV fused with PhoA/LacZ) was used as a template for the PCR in conjunction with the primers P154AF and P154AR. Upon confirmation, the mutant construct, termed pNV1997, was then transformed into SFL1616 to test for GtrIV function. Slide agglutination tests with Type IV antisera were positive. Thus, the mutant GtrIV had retained its ability to convert serotype Y to serotype 4a. Therefore, this suggests that although both GtrIc and GtrIV are structurally similar, the functions of their re-entrant loops are unique to each Gtr. As a result of this, different residues may be involved for the function of their respective re-entrant loops.

5.6 Discussion

Bioinformatics was of limited use in finding conserved residues between the Gtrs, due to the varying lengths of the Gtrs and the low homology between them. Another potential problem with bioinformatics analysis was the fact that the topology of GtrIV is different from the other Gtrs as it had a short cytoplasmic C-terminal end as compared to the long periplasmic C-terminal ends in the other Gtrs, thus giving rise to the hypothesis that the large periplasmic loop No. 6 of GtrIV is critical for function. Therefore, only the negatively charged residues (aspartic acid and glutamic acid) located throughout GtrIV were targeted as they have been found to be critical in other Gtrs (Lehane *et al*, 2005, Korres & Verma, 2006).

A total of 22 site-directed mutagenesis experiments were conducted. All of them involved mutation of the target residues to alanine (no charge). This was important as replacing a negatively charged residue with a positively charged residue may affect the overall charge of the region and effect protein function (Ulmschneider & Sansom, 2001; Johansson & Lindahl, 2006).

Therefore, subsequent loss of GtrIV function would be due to the change in charge rather than the loss of the targeted amino acid. By replacing each charged amino acid to alanine, it allows the effect of their charged side chains on GtrIV to be observed. The first 11 mutagenesis experiments targeted all acidic residues (aspartic acid and glutamic acid) located within periplasmic loops No. 2 and No. 6. The acidic residues D49, E58 and E67 located in the periplasmic loop No. 2 of GtrIV were targeted based on the belief that these negatively charged residues might be important in catalysis for the transfer and attachment of the glucosyl residue to the O-antigen as previously documented in glycosyltranferases (Breton and Imberty, 1996). Also, conserved aspartic (D) or glutamic (E) acid residues have been shown to reside within the active sides of enzymes such as lactose synthase (Gal-T1) as well as the *Bacillus subtilis* glycosyltransferase SspA (Charnrock & Davies, 1999; Ramakrishnan & Qasba, 2001). Critical aspartic acid (D) residues have also been discovered in the *Sinorhizobium meliloti* glucosyltransferase ExoM (Garinot-Schneider *et al.,* 2000). To add weight to the notion that negatively charged residues play important roles in glycosyltransferases, in recent times, critical aspartic acid residues were also discovered in glycosyltrasnferase AglD of *Haloferax volcanii* (Kaminski & Eichler, 2010). In addition, the discovery of a critical aspartic acid residue (D43) and its adjacent glutamic acid residue (E42) in GtrV, coupled with the discovery of a critical glutamic acid residue (E40) in GtrII (Lehane *et al.*, 2005), seemed to point to a conserved element in the mechanisms of action of the Gtrs. However, all acidic residues in GtrIV loop No. 2 along with the rest of the single mutations that are located in loop No. 6 were found to be non-critical. Because of the major structural differences between GtrIV and the other Gtrs, this result was not surprising as GtrIV may have a different mode of action compared to the other Gtrs. Therefore, it is possible that the acidic residues within GtrIV loop No. 2 work together in tandem

with each other. This can be investigated by cumulatively knocking out each acidic residue present in loop No. 2.

Loop No. 6 of GtrIV was found to contain 15 acidic (negatively charged) residues that are spread throughout the loop. Korres & Verma (2006) have suggested that the acidic residue D380 in GtrV is important for GtrV function because it is believed to interact with the O-antigen to stabilise it so that the glucosyl residue can come in contact with the correct rhamnose. Based on this hypothesis, any one or more of the 15 acidic residues may perform this stabilising function in GtrIV. There is also a possibility that these acidic residues compensate for the loss of the other acidic residues. This could account for the acidic residues DE261, ED283 and ED326 being non-critical. This meant that the other acidic residues could be working in tandem with the already targeted residues. The loop deletion experiments carried out in Chapter 4 helped to narrow down the search for potential residues to 19 amino acids. Of these 19 amino acids, the further deletion experiments adjusted that focus to 9 residues. All the residues, when mutated individually (C263, I264, G265, V266 and W269) did not abolish GtrIV function. However, when the acidic residues within this region were mutated in tandem with each other, it was found that the DE261 + D267 triple mutant was non-functional. The membrane protein westerns revealed that this triple mutant was localised in the bacterial membrane. As such, this result strongly suggests that the residues D261, E262 and D267 are part of a catalytic site located on GtrIV loop No. 6. Three acidic residues within GtrV (E42, D43 and D380) were found to be critical for its function (Korres & Verma, 2006). Of these three residues, E42 and D43 are located in loop No. 2 while D380 is located within the periplasmic C-terminal tail. It has been suggested that they may be involved in the formation of the active site and be responsible for the

specific α1, 3 linkage that gives rise to serotype 5a, while the re-entrant loop with specific rotational movements brought about by specific amino acid interactions is able to facilitate the addition of the glucosyl residue to the correct rhamnose of the O-antigen (Korres & Verma, 2006). In the case of GtrIV, residues D261, E262 and D267 may be responsible for the specific α1, 6 linkage that gives rise to serotype 4a. Indeed, this would explain why the all the acidic residues that were targeted in the periplasmic loop No. 2 did not have an effect of protein function unlike other members of the $Gtr_{(type)}$ family. In addition, this could also be the reason why GtrIV has this unique feature of the paired negatively charged residues occurring within periplasmic loop No. 6.

Apart from periplasmic loop No. 2 and No. 6, the re-entrant loop was another source of interest, albeit a brief one. Previous studies on GtrIc showed that P194 was critical for its function (Wang A, Honours Thesis 2011). It was hypothesised that it may be part of a "hinge" mechanism that allows loop No. 2 and loop No. 10 of GtrIc to interact with each other. With this in mind, we sought to test the function of P154 of GtrIV as it may be carrying out a similar role in this protein. This was the only proline residue within the re-entrant loop, similar to GtrIc (P194). Studies have indicated that proline residues act as a structural disrupter of secondary protein structures such as alpha helices and beta sheets (Williamson, 1994). In addition, proline residues have been considered a dominant factor for the stability of thermophile enzymes by improving the potential of hydrophobic interactions (Suzuki *et al.*, 1991). When these facts are considered, it hinted that P154 may be vital for the stability of GtrIV when it undergoes the conformational change during its catalytic activity. However, when P154 was mutated to alanine, the resulting mutant was still functional thus further underlining the fact that though GtrIV and

GtrIc are close structural homologues, each protein has evolved such that they have unique modes of action and their re-entrant loops may play different roles with regard to O-antigen glucosylation.

5.7 Conclusion

A total of 20 negatively charged amino acids that occur throughout GtrIV were mutated to alanine to investigate their role in GtrIV function. When mutated individually, they did not have an effect on protein function. However, three of these acidic residues D261, E262 and D267, when mutated in tandem, were found to abolish function of GtrIV. They are located in the periplasmic loop No, 6 and are thought to be involved in the interaction with the donor and acceptor substrates that help confer serotype specificity in O-antigen glucosylation. A proline residue (P154) located within the re-entrant loop was found to be non-critical. This study has provided a strong hint that the unique occurrence of the paired acidic residues located within GtrIV loop No. 6 may be working in tandem with the other acidic residues within their vicinity to form a potential catalytic site(s). Identifying these sites will be crucial to understanding the mode of action this topologically unique enzyme.

Chapter 6

Investigating the genetic diversity of wildtype serotype 4 strains

6.1 Introduction

There are very few biochemical properties that distinguish *Shigella* from enteroinvasive *E. coli* (EIEC), which are also a major cause of dysentery (Yang *et al.,* 2005). Indeed, some O-antigens associated with EIEC are identical to those found in *Shigella* spp (Cheasty and Rowe, 1983). It is also well documented that many plasmid-associated virulence determinants are common in both EIEC and *Shigella* (Lan *et al.,* 2001). Yang *et al* (2005) have reported that the *Shigella* chromosomes share most of their genes with that of *E. coli* K-12 strain MG1655, but there is extensive diversity of putative virulence genes, mostly acquired via bacteriophage-mediated lateral gene transfer (Yang *et al.,* 2005). Conversion evolution involving gain and loss of functions, formation of pseudogenes, insertion sequence (IS) mediated DNA rearrangements and bacteriophage-mediated gene acquisition has seen *Shigella* spp. evolve to become highly specific human pathogens with variable epidemiological and pathological features (Yang *et al.*, 2005).

The association between O-antigen modification and temperate bacteriophages in *S. flexneri* has been known for many years. Although the bacteriophages characterized to date are morphologically diverse, in all cases, the O-antigen modification genes are located immediately downstream of the phage *attP* site, which is preceded by the *int* and *xis* genes in the bacterial chromosome (Verma *et al.,* 1993; Bastin *et al.,* 1997; Huan *et al.,* 1997 a & b; Mavris *et al.,* 1997; Adhikari *et al.,* 1999; Guan *et al.,* 1999; Allison & Verma, 2000). Guan and Verma (1998) demonstrated the role played by bacteriophage mediated serotype conversion by successfully inserting a serotype conversion gene cluster into the chromosome of SFL124 through the use of a bacteriophage-based site-specific integration system. By cloning an integrase gene (*int*), an attachment site (*attP*) and a glucosyl transfer gene cluster from bacteriophage SfX into a suicide

vector, and subsequently introducing it into SFL124, they obtained a *S. flexneri* strain that expressed serotype X somatic antigen specificity (Guan & Verma, 1998). Recently, Sun *et al.,* (2011) sought to reveal possible roles played by the serotype-converting phages in the emergence of new serotypes in nature by engineering a novel serotype, named serotype 1d through the infection of *S. flexneri* serotype X strains with a SfI phage or by sequential infection of serotype Y strain with SfX and SfI.

Apart from bacteriophage mediated genetic transfer, the *Shigella* chromosome contains numerous IS elements and markers of genomic rearrangements (Buchrieser *et al.,* 2000; Jin *et al.,* 2005). IS elements are ubiquitous in bacterial genomes and important factors in evolution (Badia *et al.,* 2006). The IS insertion can cause gene inactivation, activate cryptic genes or alter the expression of adjacent genes. Roberts *et al.* (2005) reported that the presence of an IS element interrupted and inverted serotype 2a serotype-conversion locus within the chromosome of SFL124 (serotype Y), thus indicating that this strain was derived from a serotype 2a background.

Adams *et al* (2001) have described a glucosylation cassette that generates the serotype 4a in *S. flexneri* in *E. coli* K12. Even though serotype 4a in *S. flexneri* is thought to be brought about by the temperate bacteriophage SfIV, its isolation and characterisation has proved elusive. Thus far, I have managed to experimentally establish the topology of GtrIV. At the same time, site-directed mutagenesis and loop deletions have allowed us locate a potential active site within this protein. However, at the present time, not much is known about the serotype 4 strains that are circulating in the wild. Therefore, I sought to investigate the genetic diversity of the wildtype serotype 4 strains (4a and 4b) that have been isolated from developing countries such as Vietnam and Bangladesh and also in rural parts of Japan. The main purpose of this study was to gain an insight into the evolution of the wildtype serotype 4 strains circulating within these regions. In short, I wanted to determine if

there was any genetic variation of the *gtrIV* gene amongst the different isolates. By doing so, it would enable me to carry out comparative analysis to establish the level of genetic variation of *gtrIV* occurring within each region. At the same time, I also wanted to determine whether the level of diversity (if any) was region specific or common to all the wildtype strains regardless of where they were isolated from. Another key insight to be gained would be the fact that if *gtrIV* mutations are seen to occur within each strain, without affecting GtrIV function, it would help us narrow down our search for residues critical for the protein's function. Identification of genetically variant serotype 4 strains may also provide us with potential candidates for the isolation and characterization of the cryptic SfIV.

6.2 Confirmation of serotype 4 strains through slide agglutination

Currently in our laboratory strain library, out of the two hundred odd strains isolated from Vietnam, Bangladesh and Japan, 28 of them were typed as serotype 4a, 4b or just serotype 4. These strains of interest were streaked out from their glycerol stocks onto LB agar plates. A series of slide agglutination tests were carried out on the strains to confirm their serotypes. Table 6.1 lists the 28 different strains and their slide agglutination results against Type IV antisera (Denka Seiken). A total of 21 strains displayed agglutination to Type IV antisera. The remaining 7 strains (SFL1304, SFL1328, SFL2183, SFL2187, SFL2193, SFL2234 and SFL2265) did not agglutinate with Type IV antisera. As typical serotype 4a strain agglutinates to both Type IV and Group 3, 4 antisera, slide agglutination assays were also carried out against Group 3, 4 antisera. From Table 6.1, it can be seen that 8 out of the 12 strains tested as Type IV positive strains isolated in Vietnam also agglutinated with Group 3, 4 antisera. In contrast, SFL1714, SFL1767, SFL1768 and SFL1769 (all isolated from Vietnam) did not agglutinate with Group 3, 4 antisera. Similarly, all the Bangladeshi strains except SFL1523 also failed to agglutinate with Group 3, 4 antisera. As for the Japanese strains, SFL1294

Table 6.1: Summary of slide agglutination, PCR and Southern blot results of wildtype *S.flexneri* strains originally typed as serotype 4a, 4b or 4 against the various antisera

Strain	Origin	Serotype	Type IV	Group 3,4	Group 6	Group 7,8	Type II	Type V	PCR	Southern Blot
SFL1254	Control	4a	+	+	-	-	-	-	+	+
SFL1255		4b	+	-	+	-	-	-	+	+
SFL1294	Japan	4a	+	+	-	-	-	-	+	+
SFL1295		4a	+	+	-	-	-	-	+	+
SFL1304		4a	-	+	-	-	-	-	-	-
SFL1305		4b	+	-	-	-	-	-	+	+
SFL1314		4a	+	-	-	-	-	-	+	+
SFL1328		4a	-	+	-	-	-	-	-	-
SFL1522	Bangladesh	4	+	-	-	-	-	-	+	+
SFL1523		4	+	+	-	-	-	-	+	+
SFL1524		4	+	-	-	-	-	-	+	+
SFL1525		4	+	-	-	-	-	-	+	+
SFL1526		4	+	-	-	-	-	-	+	+
SFL1714	Vietnam	4a	+	-	-	-	-	-	+	+
SFL1758		4	+	+	-	-	-	-	+	+
SFL1767		4a	+	-	-	-	-	-	+	+
SFL1768		4a	+	-	-	-	-	-	+	+
SFL1769		4a	+	-	-	-	-	-	+	+
SFL2175		4a	+	+	-	-	-	-	+	+
SFL2176		4a	+	+	-	-	-	-	+	+
SFL2178		4a	+	+	-	-	-	-	+	+
SFL2179		4a	+	+	-	-	-	-	+	+
SFL2180		4a	+	+	-	-	-	-	+	+
SFL2183		4	-	+	-	-	-	-	-	-
SFL2187		4a	-	+	-	-	-	-	+	+
SFL2193		4a	-	+	-	-	-	-	-	-
SFL2234		4	-	-	-	-	-	-	-	-
SFL2241		4a	+	+	-	-	-	-	+	+
SFL2265		4a	-	-	-	-	-	-	-	-
SFL2271		4a	+	+	-	-	-	-	+	+

+ depicts a positive agglutination result
– depicts a negative agglutination result

and SFL1295 agglutinated with Group 3, 4 while SFL1305 and SFL1314 did not. LPS western blots were carried out against Group 3, 4 antisera and confirmed the slide agglutination results of those particular strains (results not shown). As some of the strains were originally typed as serotype 4, this meant that they could have also been serotype 4b strains, which are characterized by agglutination to Type IV and Group 6 antisera. At the same time, they do not agglutinate with Group 3, 4 antisera. This meant that all the strains that agglutinated with Type IV antisera but not Group 3, 4 antisera might be serotype 4b. As a result of this, all the strains were tested against Group 6 antisera and were found to be non-reactive.

Failure to agglutinate to Group 3, 4 antisera suggests that one of rhamnose sugars of the O-antigen may be carrying another modification. Therefore, all the Type IV positive strains that did not agglutinate with Group 3, 4 were tested against the Group 7, 8, Type II and Type V antisera. Type I antisera was not used as the *N*-acetylglucosamine unit is already modified with a glucosyl residue that gives rise to Type IV modification. From table 6.1, it can be seen that all the strains were negative for agglutination against the additional antisera. If agglutination had occurred against any of these antisera, for example Type II, this meant that the O-antigen of that particular strain had glucosyl groups attached to the *N*-acetylglucosamine (Type IV modification) as well as rhamnose I (type II modification). However, it is also possible that the presence of additional modification together with Type IV O-antigen modification would have prevented the other serotype specific antisera from recognising it. In other words, although the Type IV antibodies were able to recognize the Type IV modification, the O-antigen could not be recognised by the other antisera even if their respective modifications were present due to the glucosyl group attached to the N-acetylglucosamine. As such, a negative result for the slide agglutination would not necessarily mean that other O-antigen modifications are not present. The results in table 6.1 show that 2 strains

(SFL2234 and SFL2265) were negative to all serotypes and thus, were tested against Poly B antisera to confirm that they are *S. flexneri* to begin with and not some other errant strains caused by contamination on the agar plates (as they were grown on plain LB agar plates without any antibiotic selection markers).

6.3 Analysis of *gtrIV* sequences obtained from the wildtype strains

Upon confirmation of the presence of GtrIV in most of the wildtype strains, the next aim of this study was to analyse the nucleotide sequences of *gtrIV* within all these strains and compare them with the known *gtrIV* sequence. This would help us identify any possible mutations that may be present within the strains. Such analysis would help in identifying more critical residues. Hence, the 8 strains that did not agglutinate to Type IV antisera were of particular interest as successful amplification of *gtrIV* within these strains would indicate that the gene may possess critical mutations that abolished GtrIV function. The main strategy to obtain the gene sequences involved the use of colony PCR. After all the strains were streaked out onto LB agar, DNA was extracted as described in Section 2.3.2.1. Amplification of *gtrIV* was carried out with the help of the primers pGEXBamHIF and pGEXXhoIR. From the resulting colony PCR, *gtrIV* amplification was seen in all strains except SFL1304, SFL1328, SFL2183, SFL2193, SFL2234 and SFL2265 (Figure 6.1). Since these strains had negative colony PCR results and failed to agglutinate against Type IV antisera, it can be concluded that they to do not have *gtrIV*. The PCR products of the other 22 strains were excised from 0.7% agarose gel, purified and sequenced. Following this, the nucleotide sequence of each strain was aligned (using clustalW) with the *gtrIV* nucleotide sequence obtained from control wildtype serotype 4a (NCTC 8296). Similarly, the nucleotide sequences from each strain were translated into amino acid sequences and aligned against the GtrIV sequence. The results for each strain are summarised in Table 6.2. It was found that most of the strains contained 2 or

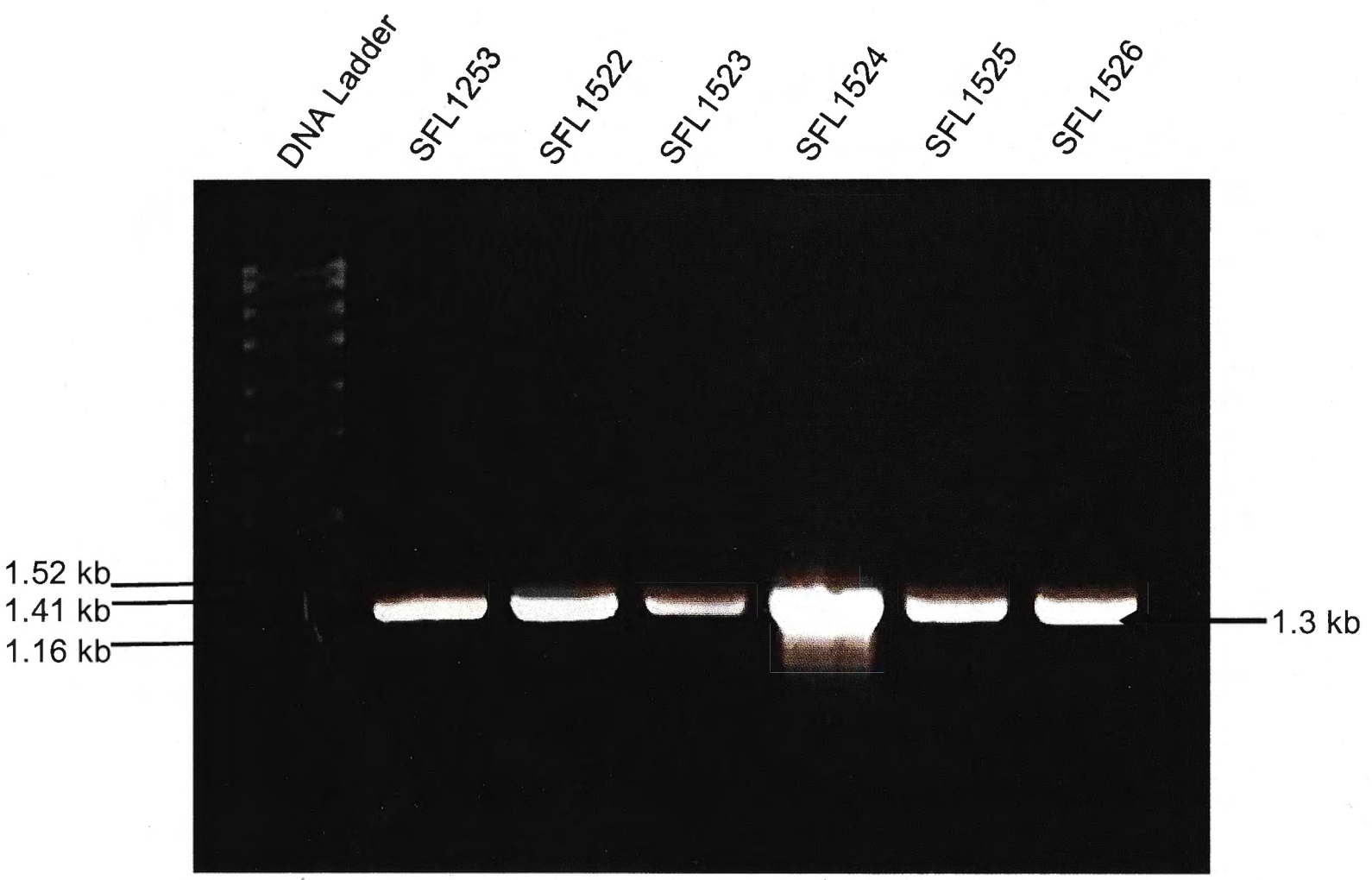


Figure 6.1: Example of gel electrophoresis results of colony PCR carried out on the wildtype serotype 4 strains. Colony PCR was carried out on the 29 type 4 strains with the help of *gtrIV* specific primers. *gtrIV* amplification was seen on 21 of those strains. The gel photo above depicts positive amplification of *gtrIV* (1.3 kb) from all 5 strains isolated from Bangladesh. SFL1253 (*S. flexneri* serotype 4a strain NCTC 8296) was used as the positive control.

Table 6.2: List of point mutations discovered within the various wildtype strains

Nucleotide Mutation	Amino acid Mutation	Strains	Location in GtrIV
ctt → gtt	L92 → V	SFL1305, SFL1526	Loop No.3
gat → gag	D146 → E	SFL1305, SFL1314, SFL1526, SFL1714, SFL1758, SFL1768, SFL2178, SFL2241	Re-entrant loop
gta → gga	V155 → G	SFL2187	Re-entrant loop
ata → gta	I159 → V	SFL2187	Re-entrant loop
aag → agg	K170 → R	SFL2187	Re-entrant loop
aat → tat	N231 → Y	SFL1305, SFL1314, SFL1522, SFL1524, SFL1526, SFL1769, SFL2178	Loop No. 6
gaa → aaa	E254 → K	SFL1314, SFL2241	Loop No. 6
gat → aaa	E283 → K	SFL1522, SFL1769	Loop No. 6
gga → aga	G434 → R	SFL1523	TM VIII
No mutations		SFL1294, SFL1295, SFL1525, SFL1767, SFL2175, SFL2176, SFL2179, SFL2180, SFL2271	

All point mutations are highlighted in red.
All Japan strains are in blue font.
All Bangladeshi strains are in brown font
All Vietnamese strains are in black font
TM VIII= Transmembrane helix VIII

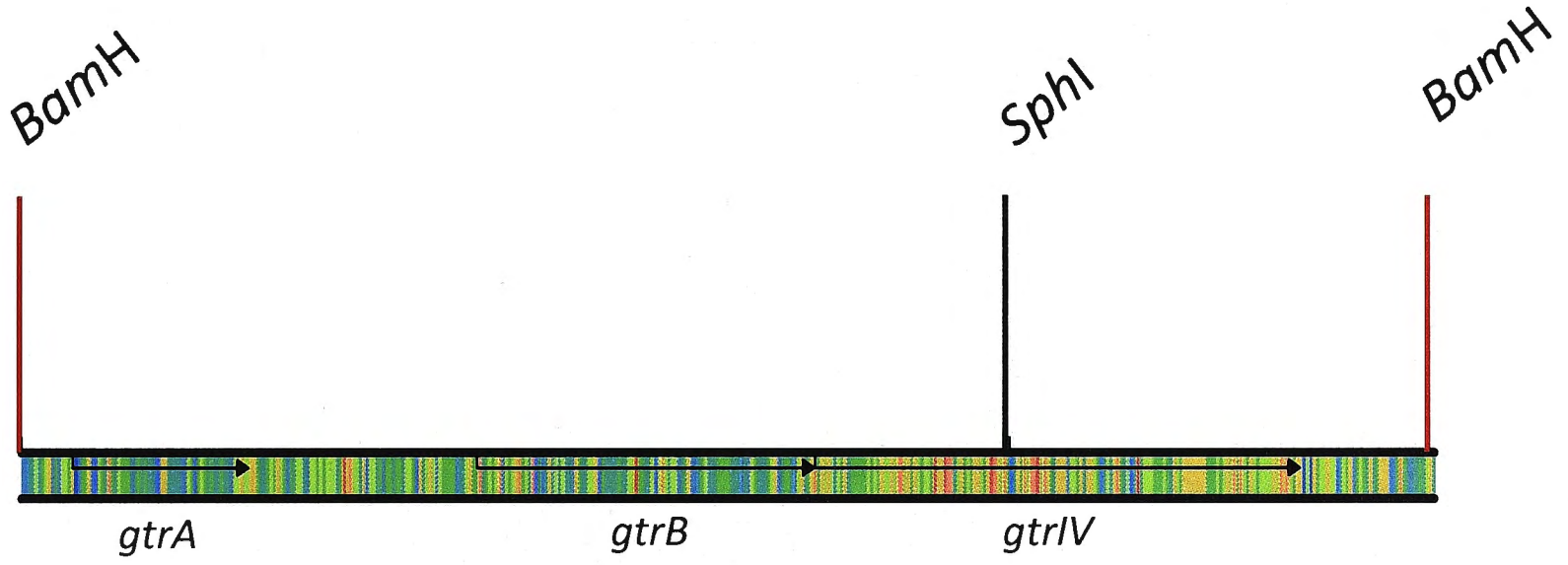
more point mutations that altered the amino acid composition of the resulting GtrIV. At the same time, there were a number of strains that bore no mutations and lined up perfectly with the Genbank sequence. It was also seen that there were no unique mutations found in any of the strains, with mutations being common amongst two or more strains. From the sequencing results, no distinct pattern of mutations was observed as the common mutations were not confined to strains isolated from the same region.

With regards to SFL2187 (did not agglutinate with Type IV antisera), it was found to contain 3 unique mutations. All three mutations were found to occur within the re-entrant loop of GtrIV. As a result of this, a further test was required to prove that the mutations were indeed responsible for SFL2187's loss of GtrIV function, thus highlighting the importance of the re-entrant loop. pBCSK was blunt-end digested with *Sma*I and the purified PCR product from SFL2187 was ligated with the digested vector and transformed into XL1-blue cells. With the help of blue-white screening, 5 colourless (white) colonies were picked and sequenced to confirm that *gtrIV* of SFL2187 was inserted in the correct orientation. Following this, the plasmid was extracted from this clone and transformed into SFL1616. This strain already contains *gtrA* and *gtrB*. Slide agglutination test using Type IV antisera was carried out and agglutination was evident. Thus, suggested that the lack of Type IV O-antigen modification witnessed in SFL2187 could have been due to an erroneous *gtrA* or *gtrB* rather than *gtrIV*.

6.4 Southern hybridization of the serotype 4 strains

Although a serotype-converting phage (the cryptic SfIV) could not be isolated from *S. flexneri* serotype 4a strains, Adams *et al* (2001) managed to isolate and characterise the O-antigen modification genes from the *S. flexneri* serotype 4a strain NCTC 8296. The 3.8kb chromosomal

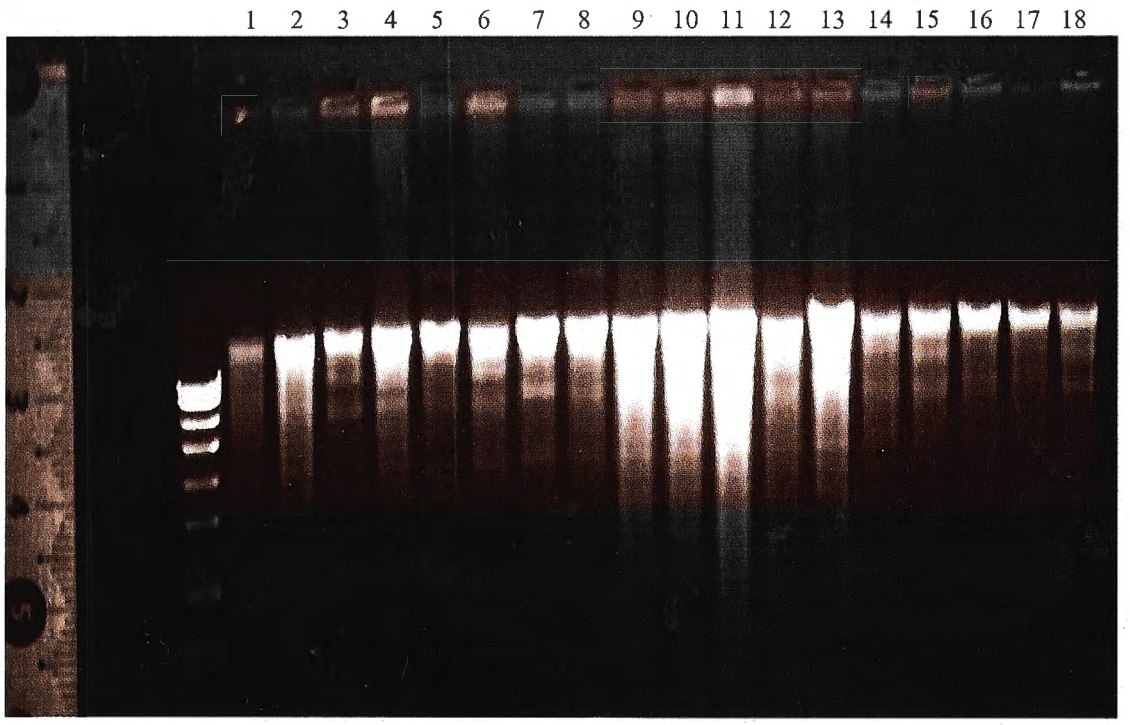
fragment from NCTC 8296 contained all three serotype conversion genes. It was isolated through restriction digestion of the chromosomal DNA with *Bam*HI (Figure 6.2). Until now however, not much is known about the rest of the *S. flexneri* 4a genome and its elusive phage SfIV. Using NCTC 8296 (SFL1253) as a control, I sought to carry out a set of Southern hybridization experiments to determine if this fragment was of the same size in the other strains. The results from this would serve two purposes. Firstly, a difference in size would highlight the possibility of differential gene arrangement occurring within that particular strain. In parallel, it would also confirm the slide agglutination results and whether those negative strains actually contained *gtrIV* to begin with. To achieve this, genomic DNA was extracted from all the 28 strains (see Section 2.2.3). Once their yield was confirmed via agarose gel electrophoresis, they were digested with *Bam*HI. The digested genomic DNA was then run on gel again and transferred onto a nitrocellulose membrane. The probe used was *gtrIV* amplified from pNV1473 (Exo III template). A total of ~1µg of *gtrIV* was labelled with the DIG easyHybrid probe. Once the probe had been prepared (see Section 2.16.2), the Southern blots were carried out as described in Section 2.16. The *Bam*HI digests and the subsequent Southern blot results are shown in Figures 6.3 and 6.4. As the loading wells had collapsed, the results of SFL2271 are not shown in these two figures. The Southern results for the other samples, however, yielded results that were mostly expected. Single bands (3.8 kb in size) were seen for most of the strains that agglutinated against Type IV antisera while no bands were visible for strains negative for Type IV agglutination. Three strains, SFL1314 (Japan), SFL1522 and SFL1526 (both Bangladesh) displayed ~10kb bands similar to the serotype 4b control rather than the 3.8kb band seen in the serotype 4a control. This suggests that the three strains could have been serotype 4b strains originally and perhaps lost function of their O-acetyltrasferase (Oac). Indeed, the lack of agglutination against Group 6 antisera (Oac specific) supports this idea. All the strains that did not display bands were subject to a repeat Southern blot to confirm the absence of *gtrIV* within those



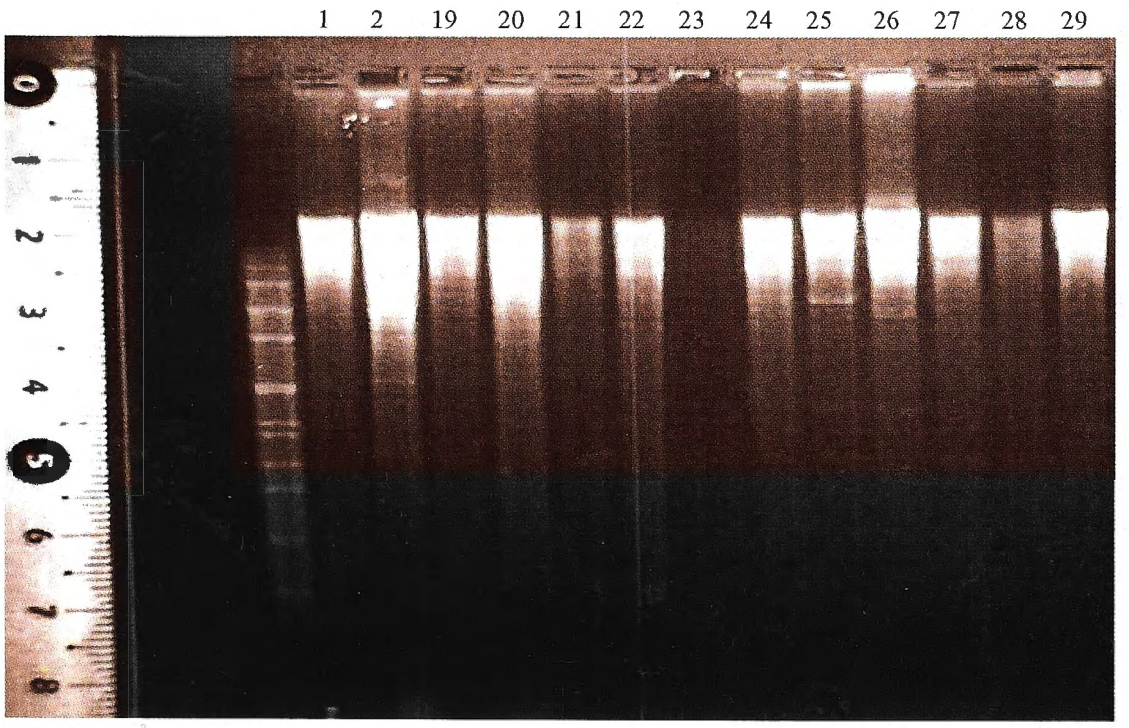
3.8 kb
gtrIV cassette

Figure 6.2: Illustration of the 3.8 kb chromosomal fragment that results from *Bam*HI digestion of the genomic DNA of NCTC 8296 (SFL1253). The fragment contains the Type IV glucosylation cassette (*gtrIV* cassette). As such, this strain will be used as a control for all the Southern hybridization experiments. The *Sph*I site is located in the middle of *gtrIV* and this digest can be used to investigate the genetic arrangement of the wildtype serotype 4a strains upstream and downstream of *gtrIV*

a)

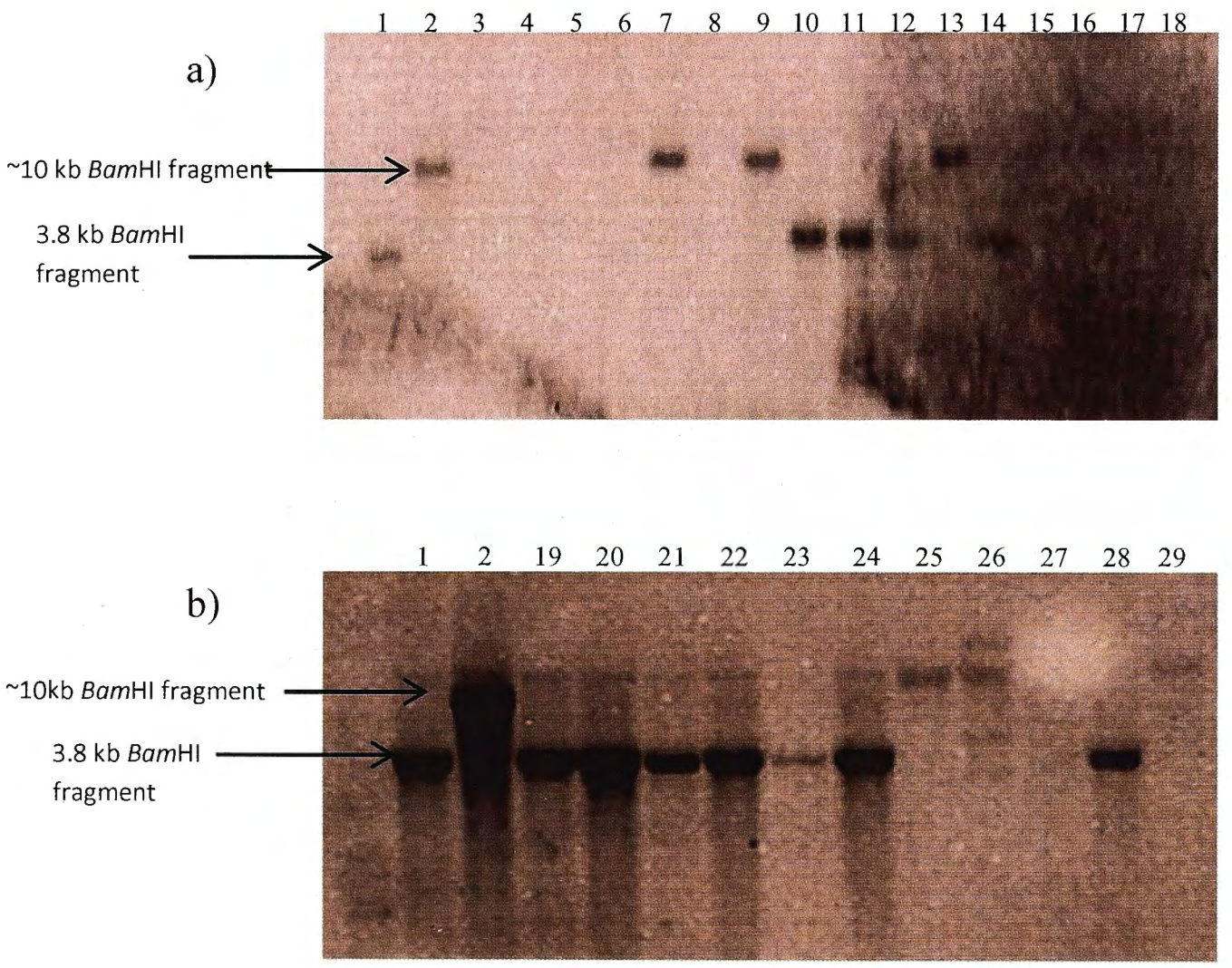


b)



1. SFL1253 (4a control)
2. SFL1255 (4b control)
3. SFL1294 (Japan)
4. SFL1295 (Japan)
5. SFL1304 (Japan)
6. SFL1305 (Japan)
7. SFL1314 (Japan)
8. SFL1328 (Japan)
9. SFL1522 (Bangladesh)
10. SFL1523 (Bangladesh)
11. SFL1524 (Bangladesh)
12. SFL1525 (Bangladesh)
13. SFL1526 (Bangladesh)
14. SFL1714 (Vietnam)
15. SFL1758 (Vietnam)
16. SFL1767 (Vietnam)
17. SFL1768 (Vietnam)
18. SFL1769 (Vietnam)
19. SFL2175 (Vietnam)
20. SFL2176 (Vietnam)
21. SFL2178 (Vietnam)
22. SFL2179 (Vietnam)
23. SFL2180 (Vietnam)
24. SFL2183 (Vietnam)
25. SFL2187 (Vietnam)
26. SFL2193 (Vietnam)
27. SFL2234 (Vietnam)
28. SFL2241 (Vietnam)
29. SFL2265 (Vietnam)

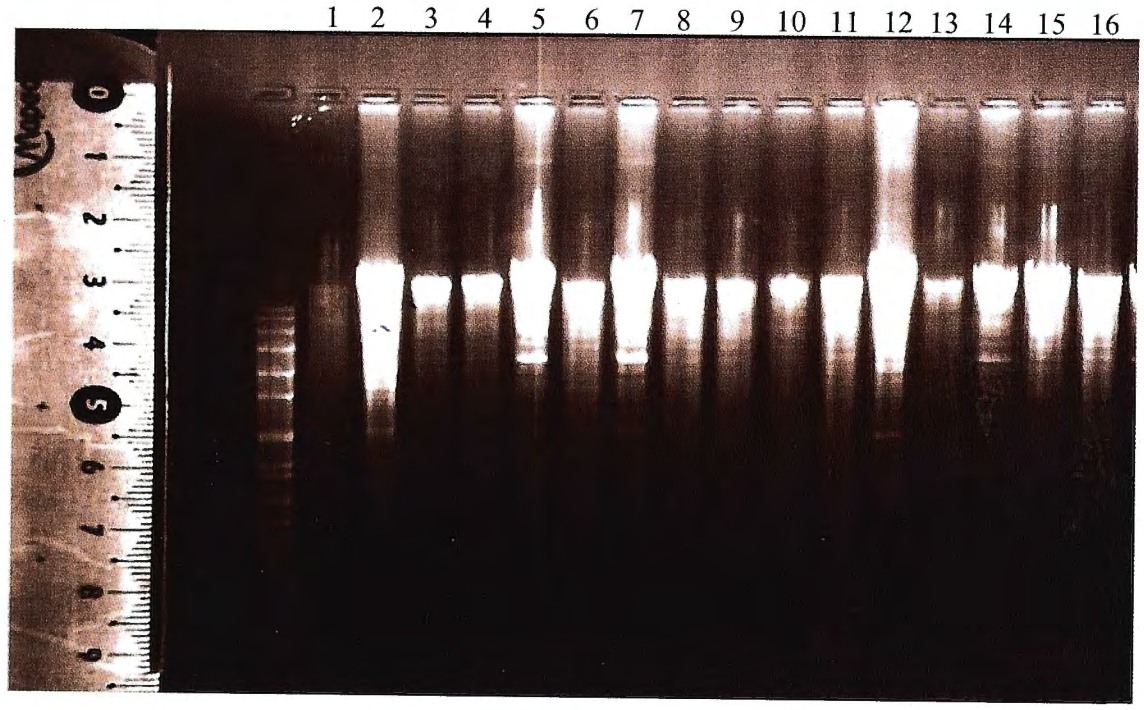
Figure 6.3: Genomic DNA of the 27 wildtype strains (excluding SFL2271) digested with *Bam*HI. All the digests were run on 0.7% agarose gels. For each digest, ~0.5 µg of genomic DNA was used. The 4a control for both gels is SFL1253 (NCTC 8296) while the 4b control used is SFL1255. The DNA for SFL2180 (lane 23) ran out of the well and as a result, the yield seen in the gel is low as compared to the other digests.



1. SFL1253 (4a control)
2. SFL1255 (4b control)
3. SFL1294 (Japan)
4. SFL1295 (Japan)
5. SFL1304 (Japan)
6. SFL1305 (Japan)
7. SFL1314 (Japan)
8. SFL1328 (Japan)
9. SFL1522 (Bangladesh)
10. SFL1523 (Bangladesh)
11. SFL1524 (Bangladesh)
12. SFL1525 (Bangladesh)
13. SFL1526 (Bangladesh)
14. SFL1714 (Vietnam)
15. SFL1758 (Vietnam)
16. SFL1767 (Vietnam)
17. SFL1768 (Vietnam)
18. SFL1769 (Vietnam)
19. SFL2175 (Vietnam)
20. SFL2176 (Vietnam)
21. SFL2178 (Vietnam)
22. SFL2179 (Vietnam)
23. SFL2180 (Vietnam)
24. SFL2187 (Vietnam)
25. SFL2183 (Vietnam)
26. SFL2193 (Vietnam)
27. SFL2234 (Vietnam)
28. SFL2241 (Vietnam)
29. SFL2265 (Vietnam)

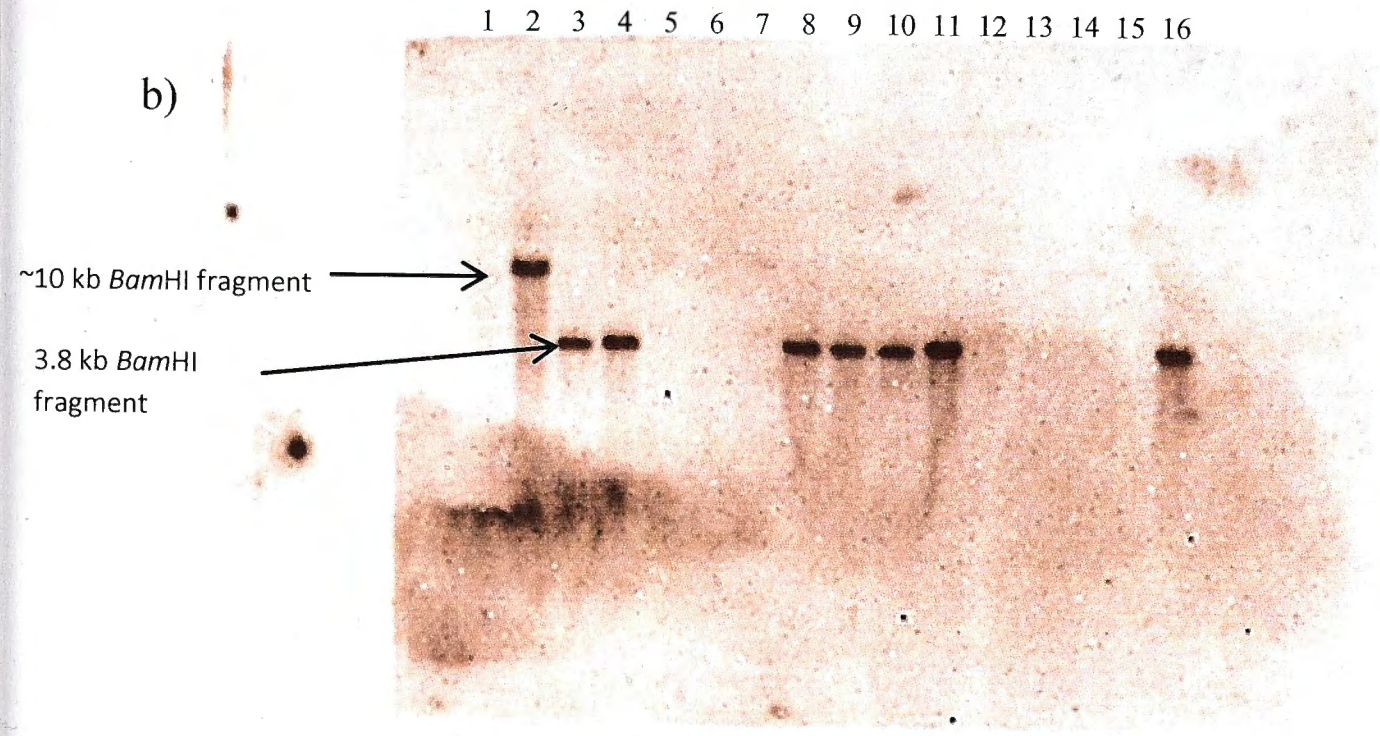
Figure 6.4: The *Bam*HI digested genomic DNA from Figure 6.3 was subjected to Southern hybridization with DIG-labelled *gtrIV* probe. a) Apart from the 4a and 4b controls, 7 strains displayed banding. Bangladeshi strains SFL1523, 1524 and 1525 had 3.8 kb bands similar to SFL1253 (4a control) along with SFL1714 (Vietnam). On the other hand, SFL1314 (Japan), SFL1522 and 1526 (both Bangladesh) have bands (~10 kb) that are similar to SFL1255 (4b control), indicating that these strains are possibly 4b strains that have lost O-acetyltransferase function. b) 7 strains, apart from the 4a and 4b control displayed 3.8 kb bands similar to SFL 1253 (4a control). The faint band seen for SFL2180 (lane 23), corresponds to the low amount of DNA seen in Figure 6.3.

a)



1. SFL1253 (4a Control)
2. SFL1255 (4b Control)
3. SFL1294 (Type IV +ve)
4. SFL1295 (Type IV +ve)
5. SFL1304 (Type IV –ve)
6. SFL1305 (Type IV –ve)
7. SFL1328 (Type IV –ve)
8. SFL1758 (Type IV +ve)
9. SFL1767 (Type IV +ve)
10. SFL1768 (Type IV +ve)
11. SFL1769 (Type IV +ve)
12. SFL2183 (Type IV –ve)
13. SFL2193 (Type IV –ve)
14. SFL2234 (Type IV –ve)
15. SFL2265 (Type IV –ve)
16. SFL2271 (Type IV +ve)

b)



strains. This repeat also included SFL2271 (Figure 6.5a and b). Another repeat was also carried out for SFL1305 (Figure 6.5c). It was seen that all the Type IV positive strains displayed 3.8kb bands under chemiluminsence, while all the Type IV negative strains were negative for hybridization to the *gtrIV* probe. The negative results confirm the absence of *gtrIV* within these Type IV negative strains.

In order to investigate the genetic arrangement of the type 4 chromosomal DNA upstream and downstream of *gtrIV,* a set of *Sph*I restriction digests were carried out as well. From Figure 6.2, it can be seen that within the *gtrIV* glucosylation cassette, there is only one *Sph*I site which lies in the middle of *gtrIV*. Thus, digesting the genomic DNA with *Sph*I would cut *gtrIV* in half. Since there is no available sequence information about the type 4 strains apart from the *gtr* cassette, carrying out this Southern would allow us to visualise if the genes upstream and downstream of this cassette in all the wildtype serotype 4 strains are arranged in a similar manner. At the same time, if one of the strains displayed a banding pattern that was larger than the rest, it might indicate the presence of the phage SfIV and that strain would be a prime candidate for phage induction. The *Sph*I digests and the subsequent Southern blots however, were inconclusive due to incomplete genomic DNA digestion (results not shown). The experiment was repeated once again but due to time constraints, the process could not be optimised.

6.5 Discussion

The work of Adams *et al* (2001) in characterising the 3.8 kb glucosylation cassette of NCTC 8296 (wildtype serotype 4a) has paved the way for the elucidation of GtrIV topology and also, for the identification of critical segments within this protein. The emergence of new *S. flexneri* serotypes, as well as the current serotype diversity witnessed in *S. flexneri* could mean that the

c)

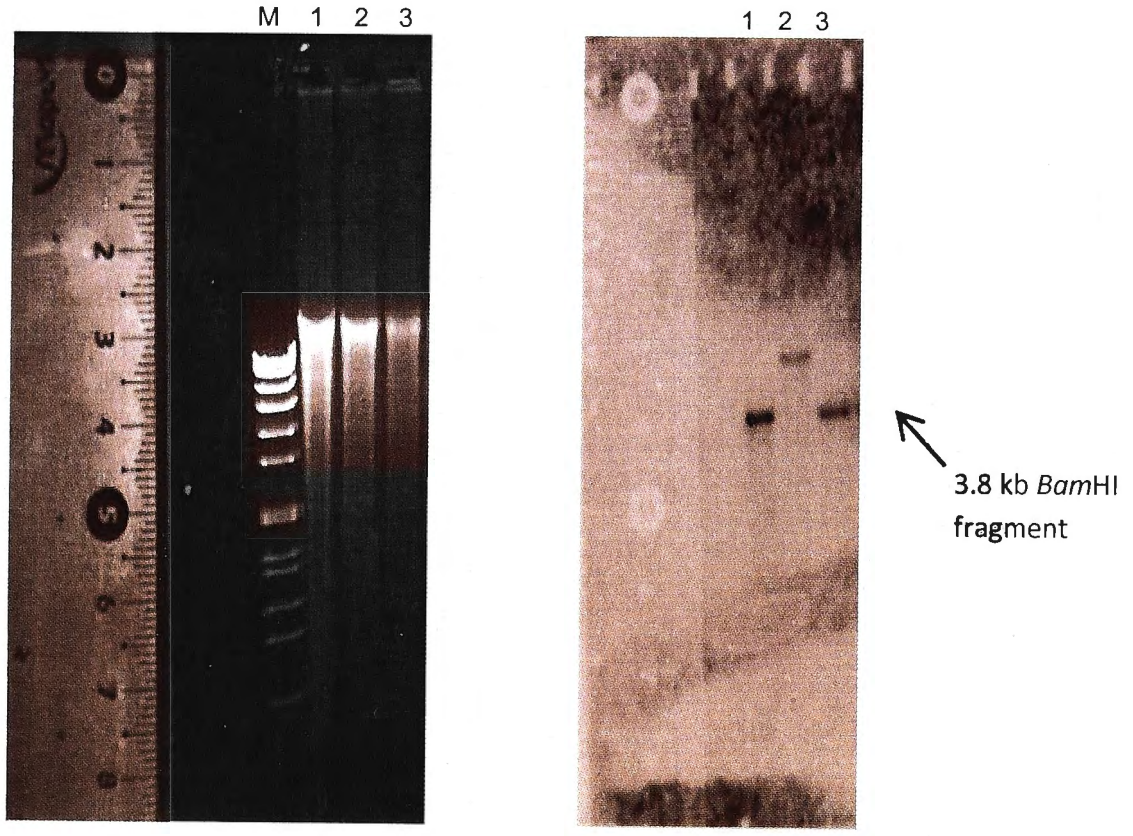


Figure 6.5: Repeat of *Bam*HI digests of genomic DNA and the subsequent Southern blot analysis. a) Genomic DNA from the strains that gave a negative result during the first round of Southern blots were digested with *Bam*HI in preparation for Southern hybridization. The 4a control SFL1253 leaked out of the well during loading of the samples and as a result, not much DNA was left. b) Southern hybridization with DIG-labelled *gtrIV* probe was carried out on the digests. This time, all strains that exhibited agglutination against type IV antisera displayed 3.8 kb bands while all strains negative for type IV agglutination did not display any banding, thus confirming the absence of *gtrIV* within those strains. As expected, no band was seen for the 4a control while the 4b control exhibited a ~10 kb band. c) The genomic DNA for SFL1305 was digested again with *Bam*HI and the Southern was repeated, yielding a 3.8 kb band.

naturally occurring serotype 4 strains may have originally been of a different serotype and eventually became serotype 4a.

Twenty-eight wildtype strains that were originally typed as serotype 4 were isolated from Bangladesh, Japan and Vietnam. To confirm their Type IV O-antigen modification, they were subjected to slide agglutination assays against Type IV antisera. Twenty one strains were positive for agglutination while 8 strains were found to be negative. Of the 21 strains that carried the Type IV O-antigen modification, 10 strains (4 from Bangladesh, 4 from Vietnam and 2 from Japan) did not display agglutination against Group 3, 4 antisera, uncharacteristic of typical serotype 4a strains. As seen in Figure 6.6, the serotype 4a O-antigen consists of a single glucosyl residue attached to the *N*-acetylglucosamine of the O-antigen repeat unit via an α1, 6 linkage. It is this modification that is recognised by the Type IV antisera. The remaining sugars that make up the rest of the repeating tetrasaccharide unit, rhamnose I, II and III, are unmodified. Thus, allowing the O-antigen to be recognised by the serotype Y (unmodified O-antigen) specific Group 3, 4 antisera. The absence of agglutination against Group 3, 4 antisera may be indicative of the presence of further O-antigen modification caused by additional glucosyltransferases (GtrI, GtrII, GtrV or GtrX) that have glucosylated the O-antigen in conjunction with GtrIV. By carrying out the slide agglutination assays against the other available antisera (type I, type II, type V, Group 7,8), the 10 strains were tested for the potential presence of any additional $Gtr_{(type)}$, making them novel serotypes. Recently in China, a study was carried out that described a *S. flexneri* isolate from China agglutinating with anti-IV type antiserum and anti-7, 8 Group antisera, indicating that this isolate possibly contained *gtrIV* and *gtrX* (Ye *et al.,* 2009). However, using monoclonal antibodies, the serotype also agglutinated with group antigen-specific monoclonal antibody MASF IV-1 but not with serotype 4 type-specific antibody

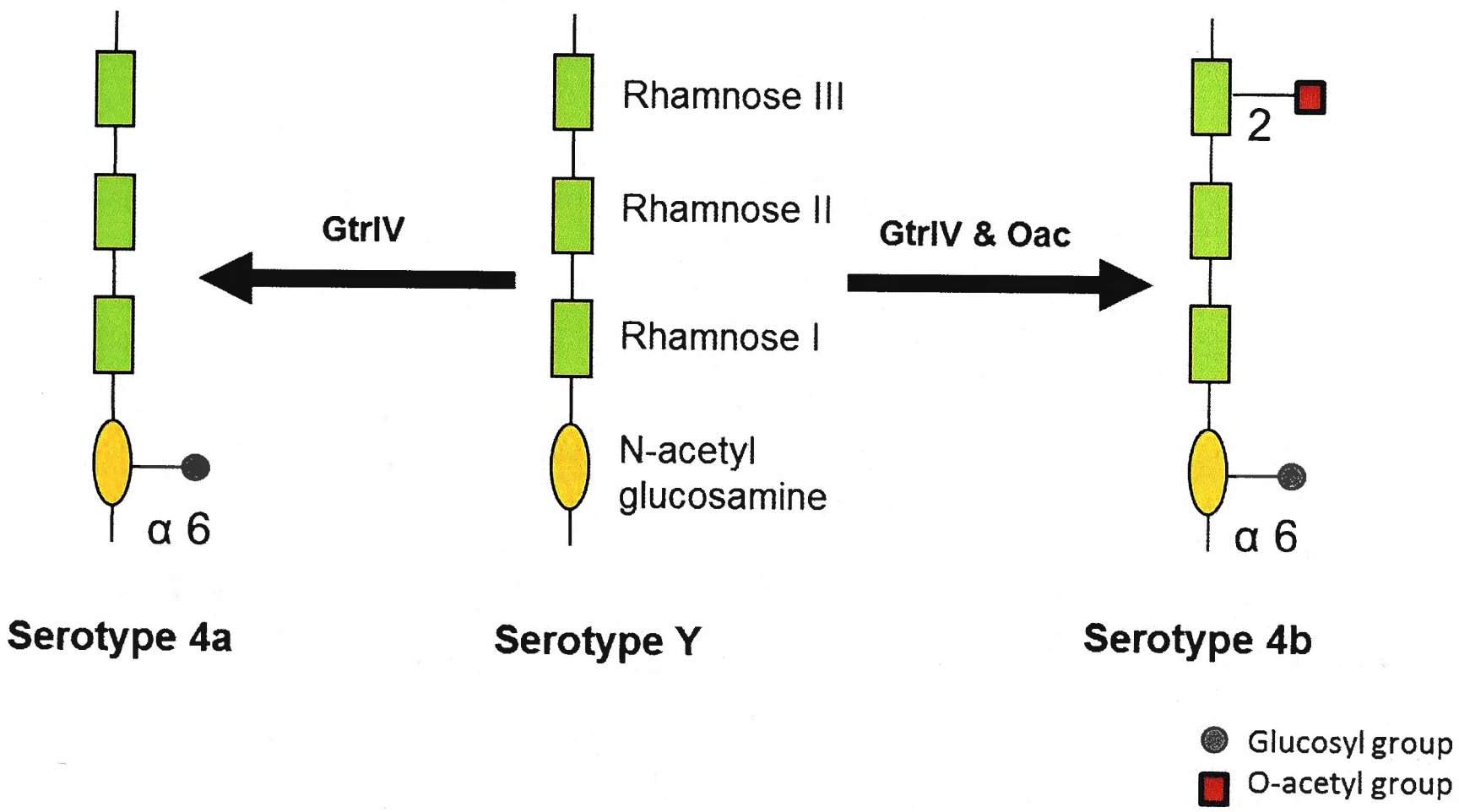


Figure 6.6: O-antigen modifications of GtrIV and Oac. GtrIV attaches a glucosyl group to N-acetylglcosamine via an α1, 6 linkage giving rise to serotype 4a. In the case of serotype 4b, in addition to the GtrIV O-antigen modification, Oac attaches an O-acetyl group to rhamnose III.

MASF IV-2 while sequencing data revealed that only *gtrX* was present in the genome (Ye *et al.,* 2009). This serotype X variant was dubbed SFxv and was claimed to be a new serotype

Although agglutination against one of the serotype-specific antisera would have immediately confirmed the presence of the $Gtr_{(type)}$ responsible, the lack of agglutination does not mean that no other $Gtr_{(type)}$ or any other unknown modification factor is present. This is because the additional O-antigen modification, in addition to Type IV modification, would have rendered the O-antigen to be undetectable by the respective serotype specific antisera. Unlike the agglutination patterns witnessed by SFxv, in the past, an uncommon serotype of *S. flexneri* type 4 was detected within Dhaka (Bangladesh) that showed a conflicting agglutination pattern by only reacting with serotype 4 specific antiserum and not with any other type- and group- specific antisera (Talukder *et al.,* 2001; Talukder *et al.,* 2002). Recently, in Korea, 3 isolates were found to have displayed a reaction with only Type IV antisera and did not bind with any of the grouping sera (Hong *et al.,* 2010). Both these cases mirror the agglutination patterns of the 10 test strains and could indicate they may in fact be a new subserotype of *S. flexneri*. Qiu *et al.* (2011) have also reported the presence of a *S. flexneri* isolate that displayed atypical agglutination patterns. This isolate was dubbed "serotype 4s" as it reacted strongly only with monovalent anti-MASF IV-type antisera but not with any group-specific antisera. Thus, indicating that it was a new subserotype of *S. flexneri* 4 (Qiu *et al.,* 2011). Based on our slide agglutination results, there is a possibility that one or all of the 10 wildtype strains isolated from Bangladesh, Japan and Vietnam may in fact be the same subserotypes of *S. flexneri* type 4. Further tests should be carried on these strains to gain confirmation. Serotyping via slide agglutination in the lab was performed using commercial *S. flexneri* antiserum (Denka Seiken, Tokyo, Japan). It would be ideal to carry out the slide agglutination assays using a panel of monoclonal antibodies against *S. flexneri* specific to all *S. flexneri* type and group antigens. Another

set of interesting results seen from Table 6.1 was the fact that 2 strains, SFL2234 and SFL2265 were not reactive to all the test antisera. They were subject agglutination against Poly B antisera to confirm that they were indeed *S. flexneri* strains. Since these strains were grown on plain LB agar plates without any antibiotic selection markers, there is a distinct possibility that a contaminant colony could have been picked out and preserved. SFL2234 tested positive for Poly B agglutination while SFL2265 was negative (results not shown). This meant that while SFL2234 was indeed a *S. flexneri* strain, SFL2265 was not. SFL2234 was also seen to be negative against Type I agglutination meaning that there was a possibility that this was a new serotype. Alternatively, SFL2234 may be a rough strain. To confirm this, LPS analysis will be required.

Apart from the crude slide agglutination experiments, the presence of *gtrIV* was confirmed by colony PCR and the resulting sequences obtained from the amplified DNA did not vary too much from the *gtrIV* nucleotide sequence obtained from the control wildtype serotype 4a strain (NCTC 8296). *gtrIV* sequence data obtained from the Type IV negative strains would have been beneficial in helping us identify critical mutations within *gtrIV*. The lack of a PCR product for the strains that did not display agglutination against Type IV antisera suggested that *gtrIV* may not have been present within these stains in the first place. This was confirmed through the Southern blots. The *gtrIV* sequences obtained from the rest of the wildtype strains were found to contain several point mutations when aligned with the GFenbank *gtrIV* sequence (Table 6.2). A total of 9 mutations were discovered among the strains. When mapped collectively, these mutations occur mainly within the re-entrant loop and loop No. 6. It is worth noting that all the mutations, except that of SFL1523 and SFL2187, were found to occur in at least two strains. These strains all shared the same nucleotide point mutations that altered the resulting amino acid sequence. In all cases, the strains that shared the same mutation did not originate from the same country. The mutations D146→E146 and

N231→Y231 were observed in 8 strains and 7 strains, respectively. D146 is located on the periplasmic face of the re-entrant loop of GtrIV while N231 occurs at the start of GtrIV loop No. 6. From the sequence data obtained, the two mutations that generated some interest were that of E254→K and E283→K. Both these mutations involved a negatively charged amino acid within loop No. 6 of GtrIV. In both cases, the strains bearing the respective mutations displayed agglutination against Type IV antisera, indicative of a functional GtrIV. This result reinforced the site directed mutagenesis results from chapter 5 which showed that when mutated individually, negatively charged amino acids within GtrIV loop No. 6 do not affect the function of GtrIV.

SFL2187 was found to have three mutations that occurred within the re-entrant loop of GtrIV. As mentioned in section 6.3, these mutations were initially thought to be critical mutations because of the strain's failure to display agglutination against Type IV antisera despite the fact that colony PCR was successful in amplifying *gtrIV* within SFL2187. However, when the amplified *gtrIV* from SFL2187 was sub-cloned into pBCSK and transformed into SFL1616 (which contains *gtrA* and *gtrB* in a separate plasmid), agglutination against Type IV antisera was restored. As such, this strongly suggests that the mutations carried by SFL2187 are not critical for GtrIV function and that this strains carries a defective *gtrA* and/or *gtrB*, thus underline the important role played by *gtrA* and *gtrB* in serotype conversion.

The Southern blot results showed that the strains negative for agglutination against Type IV antisera, apart from SFL2187, were not carrying *gtrIV* and thus, may have been mistakenly typed as serotype 4 strains or they may have lost the SfIV prophage during laboratory culture. Of the remaining 22 strains, two different band sizes were seen when the genomic DNA was digested with *Bam*HI. A 3.8kb band was witnessed in 19 of the serotype 4 strains. This band size matched up with

the expected 3.8kb fragment containing the *gtr* cassette that is flanked by *Bam*HI restriction sites. SFL1314 (Japan), SFL1522 and SFL1526 (both Bangladesh) exhibited single bands of ~10kb in size similar to the serotype 4b control (SFL1255). This suggests that these three strains could possibly be of serotype 4b origin and through time, lost the ability to O-acetylate the O-antigen. From figure 6.6, it can be seen that serotypes 4a and 4b differ by the addition of an O-acetyl group to the rhamnose III sugar of the O-antigen. This modification is mediated by O-acetyltransferase (Oac) which is encoded by the gene *oac*. O-acetylation occurs independently and does not require a glucosylation cassette, unlike GtrIV. Hence, the Southern hybridization experiments suggest that through the loss of gene function (*oac*), the serotype 4b strains have become serotype 4a. Several experiments can be carried out to confirm the presence of *oac*. One can attempt to amplify this gene through PCR and any amplification witnessed would be proof of *oac* presence. The loss of *oac* function could be attributed to deletions and pseudogenes, which are effective mechanisms that contribute to loss of function (Yang *et al.,* 2005). Such gene decay or reductive evolution is noted to be an important evolutionary mechanism for obligate intracellular pathogens such as *Mycobacterium leprae* (Cole *et al.,* 2001). The loss of gene functions for adaptations and convergent evolution has also been documented within *Shigella*. The inactivation of *ompT* (encodes for the surface protease OmpT) and *cadA* (gene encoding lysine decarboxylase) genes have increased the pathogenicity of *Shigella* (Nakata *et al.,* 1993; Sakellaris *et al.,* 2000; Lan and Reeves, 2002). In the case of *opmT*, a gene present in *E. coli* K-12, the presence of the surface protease OmpT, was found to attenuate the virulence *Shigella* by degrading the IcsA (VirG) protein (Nakata *et al.,* 1993; Sakellaris *et al.,* 2000). IcsA (VirG) is required for actin-based movement within infected cells (Goldberg *et al.,* 1993; Goldberg & Theriot, 1995).

The Southern hybridization experiments were carried out in order to visualise the 3.8 kb glucosylation cassette present within wildtype serotype 4a. Therefore, in theory, seeing as *oac* is independent of this cassette, the same sized band should have been picked up from all the strains regardless of serotype. The *oac* gene coding for the O-acetylation factor is carried by the temperate bacteriophage Sf6 of *S. flexneri* (Thanweer *et al.,* 2008). As such, this may account for the difference in size of the bands witnessed as the insertion could have occurred in the region that of the genomic DNA that contains the *BamH*I restriction site. Alternatively, as allelic variations in bacteria arise from random mutations, point mutations would have occurred in the genome such that one of the *Bam*HI restriction sites may have been destroyed. This occurrence is termed as single nucleotide polymorphisms (SNPs). SNPs are a DNA sequence variation occurring when a single nucleotide in the genome (or other shared sequence) differs between members of a species (Barreiro *et al.,* 2008). By this virtue, it is plausible to speculate that the same random mutations could have caused another *Bam*HI site to be generated much further upstream/ downstream of the *gtrIV* cassette, giving rise to a much larger band. Similarly, without further confirmation for the presence of *oac*, it may also be that the three strains are actually serotype 4s strains carry mutations on their *Bam*HI sites and as a result, generated larger *Bam*HI digest fragments.

By subjecting the genomic DNA of the wildtype strains to *Sph*I digests, with the help of Southern hybridization, it would enable visualisation the sort of genetic arrangement present upstream and downstream of the *gtrIV* cassette. By comparing the various arrangements, I might have been able to determine the strain that still may have had the cryptic bacteriophage SfIV embedded within its chromosomal DNA. Therefore, had the *Sph*I Southerns been successful, comparisons could have been made among the strains. Another way to carry out such comparisons is through genome sequencing and is described in more detail in the discussion (Section 7.3).

6.6 Conclusion

Twenty eight wildtype strains isolated from Japan, Bangladesh and Vietnam that were typed as serotype 4a, 4b or just 4 were subjected to a series of slide agglutination experiments to confirm their serotype. A number of strains had atypical agglutination patterns and thus highlighted the possibility that they are potentially new serotypes or subserotypes. Colony PCR was also carried out and the amplified DNA from the PCR reactions were sequenced and aligned to the control wildtype 4a strain SFL1253 (NCTC 8296). The sequences obtained revealed the presence of point mutations within several strains, although I was not able to identify mutations that were detrimental to the function of GtrIV. Southern blots carried out using DIG-labelled *gtrIV* revealed that two different serotype 4a strains, with the possibility of one of them being originally a 4b strain, have been circulating within these populations. A further attempt to elucidate the genetic arrangement of these strains proved to be unfruitful. Taken together, this preliminary study has highlighted the possibility of 2 serotype 4a strains currently circulating in Vietnam, Bangladesh and Japan. This study has provided valuable information for further investigation into the evolution of one of the most important human pathogens.

Chapter 7

General Discussion

The LPS of *S. flexneri* is an important virulence factor required for protection against host defences (Oaks *et al.,* 1985; Hong & Payne, 1997; Schroeder & Hilbi, 2008). It is therefore paramount that the development of a multi-valent live attenuated vaccine is dependent on the creation of a single *S. flexneri* strain with a LPS that elicits immunogenic responses for more than one serotype. To achieve this, the mechanism of O-antigen (LPS component) modification, which results in serotype conversion, must be clearly understood. O-antigen modification is mediated by glucosyltransferase genes (*gtr*) encoded by temperate serotype-converting bacteriophages. GtrIV, an integral membrane protein, is responsible for attaching a glucosyl residue to the *N*-acetylglucosamine of the O-antigen repeat unit via an α1, 6 linkage, thus mediating the conversion of serotype Y to serotype 4a.

7.1 Topological studies on GtrIV

By using a dual reporter system consisting of PhoA/LacZ we were able to solve the secondary structure of GtrIV by creating C-terminal GtrIV-PhoA/LacZ deletion fusions and GtrIV-PhoA/LacZ-GtrIV sandwich fusions. GtrIV was found to contain 8 transmembrane helices, cytoplasmic N- and C- termini, a large periplasmic N-terminal loop (loop No. 2) and large periplasmic C-terminal loop (loop No. 6) (see figure 3.2)(Nair *et al.,* 2011). The existence of a unique re-entrant loop was also shown to occur between transmembrane helices III and IV. Of all the known Gtrs, this unique topology resembles that of GtrIc the most (Ramiscal *et al.,* 2010). Though both proteins possess the large periplasmic N- and C- terminal loops, the configuration of their re-entrant loops sets them apart. This study has highlighted the fact that the re-entrant loop of GtrIV starts on the periplasmic face of transmembrane helix 3, traverses across the lipid bilayer down to the cytoplasm, before making its way back to the periplasm. In contrast, the re-entrant loop

of GtrIc starts at the cytoplasmic face of transmembrane helix IV forms a double intramembrane dipping structure and joins the cytoplasmic face of transmembrane helix V (Ramiscal *et al.,* 2010). As mentioned in Chapter 3, the existence of a re-entrant loop has previously been reported in eukaryotic glutamate transporters and bacterial potassium channel KcsA and also in glycerol and water channels (Sonders & Amara, 1996; Seal *et al.,* 2000; Grunewald *et al.,* 2002; Brocke *et al.,* 2002). In the latter two cases, the two loops depicted in these channels play a major role in permeation (Sonders & Amara, 1996; Seal *et al.,* 2000). In contrast to the pore loops of KcsA and the aquaporins, these re-entrant loops are not fixed and have been shown to undergo conformational changes in the presence of sodium (MacKinnon R, 1995; DeFelice & Blakely, 1996; Sonders & Amara, 1996).

The use of the PhoA/LacZ dual reporter system has been highly successful in determining the topology of several other membrane proteins in *Shigella flexneri* (Korres & Verma, 2004; Korres *et al.,* 2005; Lehane *et al.,* 2005; Thanweer *et al.,* 2008; Ramiscal *et al.,* 2010). This system is however, not without its limitations. Though it has allowed us to determine the orientation of the various loops and the transmembrane regions, it does not allow us to know the transmembrane boundaries and proximities between the transmembrane helices and connective hydrophilic loops. Therefore, one can continue this topological analysis with the help of cysteine scanning mutagenesis (SCAM) (Bogdanov *et al.,* 2005; Lehrer *et al.,* 2007) and thiol cross-linking (Kwaw *et al.,* 2000; Wolin & Kaback, 2000). SCAM is a method which can be used to investigate structural alterations in different functional states of membrane proteins as well as determine the size of catalytic sites and any residues within (Liapakis *et al.*, 2001). This system involves mutating all cysteine residues of the GtrIV to serine, thus creating a cysteine-less mutant (Bogdanov *et al.*, 2005). In order for this technique to be successful, this mutant must be functional, after which specific cysteine residues can

be introduced in place of residues that span the hydrophilic loops and transmembrane regions. This would create a library of clones that have single cysteines introduced to different positions along the hydrophilic loops and corresponding transmembrane helices of GtrIV. Depending on the reaction of the introduced cysteine residues to a combination of reagents such as [2-(trimethylammonium)ethyl] methanethiosulfonate bromide (MTSET) and N^{α}-(3-maleimidylpropionyl) biocytin (biotin maleimide) under different conditions (such as membrane permeability), the orientation and organisation of each transmembrane segment in its native membrane can be determined (Bogdanov *et al.*, 2005). This will allow us to determine the length of transmembrane helices and the distance between each helix. A key drawback to this system, however, is the fact that the cysteine-less (Cys-less) GtrIV mutant must be functional. The knocking out all 6 cysteine residues within GtrIV individually might not pose a problem but the cumulative burden of all cysteines being knocked simultaneously may prove to be too much to bear for GtrIV. Recently, researchers in our lab have successfully created a functional GtrV protein devoid of all 6 of its cysteine residues. This means that there is a chance that a Cys-less GtrIV mutant might also be functional.

Thiol-cross linking can also be used to investigate transmembrane helix proximity as well as hydrophilic loop proximities by introducing cysteines at specific sites along the GtrIV protein. This technique has been successfully employed in the past to determine the proximity of cytoplasmic loops and transmembrane domains of lactose permease in *E. coli* (Kwaw *et al.*, 2000; Wolin & Kaback, 2000). It also requires a functional GtrIV cysteine-less mutant. By introducing a protease Xa site to the middle of the protein, it can be cleaved into two separate fragments. Specific tag such as FLAG or c-myc can be used for detection of the protein on western immunoblots. Upon introduction of the protease Xa site and confirmation of GtrIV function, cysteines can be introduced to specific loops of GtrIV, for example if the proximities of loop No. 2 with loop No. 10 need to be

investigated, one cysteine will be introduced in loop No 2 and another in loop No 10. The membrane extracts can then be treated with different chemical cross-linking reagents that form different bonds (rigid or flexible bonds) at varying lengths (ranging from 6Å to 16Å) such as *N,N'-o*-phenylenedimaleimide (o-PDM) and 1,6-bis(maleimido)hexane (BMH) before being cleaved with factor Xa protease (Kwaw *et al.,* 2000; Wolin & Kaback, 2000). If there is a potential cross-link between the introduced cysteines with the above mentioned reagents then the protein will not be separated in two separate fragments and thus a reconstituted protein will be detected on the immunoblots with incubation using the antibody specific to the tag introduced. If a cross-link is not present then only the half of the protein which has the specific tag attached will be detectable on the immunoblots. Cross-linked samples can be confirmed by subjecting them to a reducing agent such as Dithiothreitol (DTT) which would disrupt the bond between the two cysteines and revert the detection to only the fragment of the protein which has the specific tag attached to it. As such, additional residues on the target loops will also be tested to confirm or establish potential cross-links and proximity results.

7.2 Functional analysis of GtrIV

As glucosylation is thought to occur in the periplasm, the large periplasmic loops No. 2 and No. 6 of GtrIV were prime candidates for functional studies. A series of deletion experiments were aimed at the periplasmic loops No. 2 and No. 6, as well as cytoplasmic loop No. 3. Because if its large size, loop No. 6 was then deleted in smaller segments. All three segments were chosen such that they each encompassed the paired acidic residues DE261, ED283 and ED326 as well the negatively charged residues that flank them. From the resulting partial deletions and with the help of membrane protein western blots, one of the non-functional deletion proteins, GtrIV loop No. 6 partial deletion 3, was found to be localised within the bacterial membrane in amounts similar to that

of the intact GtrIV. This was in stark contrast to the other deletion proteins which had little to no protein localised in the bacterial membrane (Nair *et al.,* 2011). As mentioned in chapter 4, the lack of protein assembly in the membrane could be attributed to a number of factors such as disruption of van der Waals forces that have affected the hydrophobicity of the deletion proteins or disruption of inter-helical interactions that are essential for stability of the tertiary protein structure (Harrington & Ben-Tal, 2009; Fiedler *et al.,* 2010). In both cases, the result would have been an unstable, misfolded protein which would have been unable to localise within the cellular membrane, thus resulting in its degradation within the cytosol. This led to the belief that the missing 19 amino acid segment of loop No.6 partial deletion 3 could contain residues that may be part of a catalytic site present in the large periplasmic loop No. 6. Further deletions were carried out and narrowed the potentially critical region to 9 amino acids (see chapter 4). From these 9 amino acids, site-directed mutagenesis revealed that only when residues D261, E262 and D267 were knocked out in tandem, did GtrIV lose its function. Confirmation of this was done through LPS and membrane protein western blots which showed that the triple mutant was non-functional yet still assembled in the cellular membrane. The strategy of sequentially deleting segments and following up with site-directed mutagenesis may prove to be a more efficient way to uncover critical residues within GtrIV. This is because of the fact that although critical acidic residues have been located within loops No. 2 of the other Gtrs, they still remain elusive within GtrIV. Therefore, it would be a good idea to sequentially delete loop No. 2 of GtrIV and carry out functional analysis similar to what was undertaken with loop No. 6. In addition, one should also knock out the acidic residues within GtrIV loop No. 2 in tandem as like those of loop No. 6, these negatively charged residues (5 in total) may either be working in tandem with each other or may be compensating for the loss of any of the other acidic residues within the loop.

LPS and membrane protein western blotting were also employed for chimera studies between GtrIc and GtrIV. Swapping of a single large periplasmic loop between GtrIc and GtrIV effectively created four chimeras. Functional analysis showed that all four chimeras were non-functional. Membrane protein westerns showed that apart from the GtrIc-GtrIVlpNo.2-GtrIc chimera, all other chimeras were present in the membrane. This result led to the assumption that interaction of both periplasmic loops are required for $Gtr_{(type)}$ proteins to function. However, double loop chimeras between GtrIc and GtrIV whereby both loops No. 2 and No. 6 of GtrIV were replaced by loops No. 2 and No. 10 of GtrIc and vice versa, were created and the resulting hybrids neither regained their native function nor gained their substitute function (Wang A, Honours Thesis 2011).

Although all the deletion proteins, chimeric proteins and mutant proteins used in this study were successfully constructed and the LPS western blots provided definitive and consistent results, one main drawback with regards to the membrane protein western blot remains the fact that the results cannot be quantified with a high degree of accuracy. As such, intense bands may not be a direct result of good protein expression instead, they may also be due to overexposure to the labelled antibody. To get around this issue, different western blotting systems such as near-infrared fluorescence (NIF) for western detection are now available (Weldon *et al.,* 2008; Mathews *et al.,* 2009). This method involves using fluorophores which are directly conjugated to the antibody probes used in western blotting (Weldon *et al.*, 2008). With the help of a new laser-based imaging system by LI-COR biosciences, the resulting fluorescence can then detected (Mathews *et al.,* 2009). Boasting sensitivity superior to chemiluminescence, NIF is also able to quantify relative protein expression in different samples (Weldon *et al.*, 2008). This would have allowed for a more thorough analysis of all the different modified proteins created. For instance, the GtrIV loop No. 6 deletion and partial deletion proteins all made their way to the membrane. However, they were present in

small quantities as compared to the intact GtrIV. As such, with the use of this system, one can quantify all the mutant/deletion proteins from future experiments and compare them to the intact GtrIV.

7.3 Genetic diversity of wildtype serotype 4 strains

A preliminary study was carried out to gain an insight into the genetic diversity of wildtype serotype 4a strains. Of the 28 strains isolated from Japan, Bangladesh and Vietnam, 10 strains displayed atypical slide agglutination patterns, suggesting that they might potentially be new serotypes or subserotypes. Up till now, the slide agglutination method against specific type and group factors has been the only means to routinely identify the serotypes of *S. flexneri*. However, as mentioned in previous chapters, an incorrect reading can occur as a result of visual assessment of the agglutination reactions leads to conflicting agglutination results being re-examined through LPS western blotting. As for molecular serotyping, up to 10 separate agglutination tests using antisera for type antigens I, II, III, IV, V, and VI and for group antigens 3;4, 7;8, and 6 and the monoclonal antibody MASF1c for serotype 1c are required (Talukder *et al.,* 2003; Stagg *et al.,* 2008; Ye *et al.,* 2010). This coupled with the high cost of antiserum kits limit its application in the laboratories of developing countries. Recently, Sun *et al* (2011) have developed a multiplex PCR assay especially for molecular serotyping of *S. flexneri* that targets the O-antigen synthesis gene *wzx* and the O-antigen modification genes *gtrI*, *gtrIC*, *gtrII*, *oac*, *gtrIV*, *gtrV*, and *gtrX*. This assay contained eight sets of specific PCRs in a single tube and can identify 14 of the 15 serotypes (the exception being serotype Xv) of *S. flexneri* in a single reaction. This method would be ideal for identifying potential new subserotypes especially with the Bangladeshi and Vietnamese strains that agglutinated against Type IV antisera but not against Group 3, 4 antisera. Other advantages to this method include the

fact that it can be utilised for high-throughput serotype identification with results being obtained in 3 to 4 hours (Sun *et al.,* 2011). Following this, the results can then be confirmed through serotyping.

Sequencing of the wildtype strains revealed the presence of point mutations within several strains, while Southern blots carried out using DIG-labelled *gtrIV* revealed two different bands for the serotype 4a strains. Although one of the bands is similar to that the serotype 4b strains, before one can confirm that these strains are indeed serotype 4b strains that lost the function of *oac*, attempts must be made to try and amplify the *oac* gene from these strains. Also, Southern blotting (with DIG-labelled *oac* as the probe) can be carried out on these strains. A better understanding of the intra-species diversity of *Shigella* requires the availability of whole genome sequences. Given that only a 3.8kb region of the whole serotype 4a genome has been sequenced, it would be useful to gather sequence information on the flanking regions of the *gtrIV* cassette. Obtaining the whole genome sequence of serotype 4a would allow us to make comparisons with the other *S. flexneri* genomes to identify any differences within them. In the past, the complete genome sequence of *S. flexneri* 5b (Sf8401) and a comparison with most of the *S. flexneri* 2a (Sf301) genome revealed differences in the pathogenicity islands and chromosomal rearrangements between different serotypes of this species (Nie *et al.,* 2006). In the same study, the authors showed that there was ~15 kb of bacteriophage SfV sequence remnants in Sf8401, though far less bacteriophage SfII sequence has remained in Sf301. This may suggest that Sf8401 arose later than Sf301. In this vein, by obtaining the whole genome sequence of serotype 4a, similar comparisons can be made to help facilitate understanding of the common biological processes required for infection and identify unique properties that may differentiate between them in respect of epidemiology and pathogenicity, even if the virulence plasmid is closely similar. Moreover, the comparison will provide some insight into how these pathogens have evolved.

7.4 GtrIV, a distinctly different glucosyltrasferase

This study has highlighted that GtrIV has a unique topology compared to the other Gtrs. Its closest structural homologue, GtrIc, is 526 amino acids long compared to the 437 amino acids of GtrIV. Bioinformatics comparisons between the two proteins revealed that although they are structurally similar, both proteins sequences only shared 4.6% identity and 8.7% similarity. Ramiscal *et al.,* (2010) stated that the distinct topological differences between GtrIc and the other Gtrs can be attributed to the fact that it catalyses a novel glucosyl to glucosyl linkage on the O-antigen that is not present in any other serotypes. In the case of GtrIV, it has a strikingly different topology to the other Gtrs even though, like the other Gtrs, it catalyses a glucosyl group directly onto one of the sugar residues that make up the O-antigen. In terms of O-antigen modification, GtrI is closest to GtrIV as it attaches a glucosyl residue to the *N*-acetylglucosamine of the O-antigen repeat unit via an α1, 4 linkage. Their topologies though, differ quite significantly as GtrI has a periplasmic C-terminal tail which is thought to attach the glucosyl group to the O-antigen in a site and linkage specific manner. Another unique GtrIV trait is the high number of acidic residues (aspartic acid and glutamic acid) present within loops No. 2 and No. 6, resulting in neither of the residues being critical on their own.

The relationship between GtrIV and the other glycosyltransferases that use sugar-nucleotide donors (GT) is insignificant. The GT-A superfamily is characterized by the presence of a DxD motif and depends on metal ions for activity (Unligil & Rini, 2000; Bourne & Henrissat, 2001). This motif is seen in loop No. 2 of GtrIV but not critical for its function. There is also no known association between GtrIV and any metal ions. On the other hand, another superfamily GT-B does not require metal ions for enzymatic activity but shows weak similarity to GtrIV (Hu *et al.,* 2003). This further underlines the enzyme's unique function and mode of action. By comparison, the enzymes from

both GT-A and B superfamilies are quite diverse from the multi-transmembrane O-antigen modifying glucosyltransferases of *S. flexneri*. Since GtrIV uses a lipid-linked donor such as UndP-β-glucose, it can be compared to such other proteins from a range of organisms (Hu *et al.,* 2003). For example, glycosyltransferases that use Dol-P-linked monosaccharides (Dol-P-Mannose, and Dol-P-Glucose) as the donor substrate show similar topological arrangements. Apart from the fact that these glycosyltransferases consist of at least 6-12 transmembrane segments, they all share a highly conserved N-terminal loop after a transmembrane segment, similar to loop No. 2 of GtrIV (which is after transemembrane helix I)(Oriol *et al.,* 2002; Baulard *et al.,* 2003).

The results obtained throughout the course of this project have pointed to the fact that GtrIV may have a mechanism of action that different from the rest of its Gtr counterparts. From the models proposed for the mechanisms of action of GtrII, GtrV and GtrIc, it is generally believed that the periplasmic loop No. 2 of the Gtrs binds with glucose-precursor and periplasmic loop No. 10 (GtrIc) or the periplasmic C-terminal tail (GtrII and GtrV) are believed to stabilise the O-antigen by interacting with the specific sugar residue that is going to be glucosylated (Lehane et al., 2005; Korres & Verma, 2006; Ramiscal *et al.,* 2010). The re-entrant loops confer flexibility to these enzymes by allowing them to bend such that both the periplasmic domains can come into close contact to transfer and attach the glucosyl group to the O-antigen. This mode of action, however, cannot be used to describe GtrIV for a number of reasons. First of all, it is believed the D/E x D/E motif in loop No. 2 is required to deprotonate the hydroxyl group on the sugar residue of the O-antigen so that a glucosyl group can be attached (Ramiscal *et al.,* 2010). This hypothesis is underpinned by the fact that D40 (GtrIc), E40 (GtrII), E42 and D43 (both GtrV) are critical for the function of their respective Gtrs. The periplasmic loop No. 2 of GtrIV contains this motif and both acidic residues D31 and D33 within this motif were found to be non-critical (Nair A, Honours

Thesis 2006) while all the other acidic residues within this loop were also non-critical. This finding, coupled with the discovery of three critical residues (when knocked out in tandem) in GtrIV loop No. 6 brings out two possible but highly speculative modes of action.

Over time, evolution could have rendered the presence of periplasmic loop No. 2 of GtrIV to be purely structural. As such, all the glucosylation work is carried out within loop No. 6 itself. In which case, D261, E262 and D267 would obtain the glucosyl group from the bactoprenol-glucosyl complex and assist in deprotonating the nucleophilic hydroxyl group of the acceptor sugar (*N*-acetylglucosamine) while a combination of the other acidic residues, in tandem with the paired residues ED283 or ED326, stabilise the O-antigen so that the glucosyl residue can come in contact with the *N*-acetylglucosamine. The UndP carrier is then flipped to the cytoplasmic leaflet of the inner membrane to be used again. This model takes into account the fact that no acidic residue was found to be critical within GtrIV loop No. 2. Alternatively, the acquisition of a glucosyl group and the subsequent deprotonation of the hydroxyl group of the acceptor sugar residue could be carried out by anyone of the 5 acidic residues within loop No. 2. In this scenario, anyone of the 5 acidic residues may carry out this role and therefore, act to compensate for the loss of the other acidic residues or they may be acting in combination of each other. D261, E262 and D267 from loop No. 6 would then bind to the O-antigen. The re-entrant loop then acts as a 'hinge' to bend the protein allowing the O-antigen to be stabilised and presented in a manner that allows the glucosyl group to be attached via an α1, 6 linkage. The UndP carrier is then flipped back into the cytoplasm.

In order to prove or disprove these hypothetical models, one must knock-out all the acidic residues in GtrIV loop No. 2 in combination. In addition, further deletion experiments should be

carried out on the other loop No. 6 partial deletions (1 and 2). This would help generate more data to help clarify the mode of action of this unique Gtr.

7.5 Moving forward: Future work on GtrIV

Apart from topological and functional studies on GtrIV, not much else is known of this integral membrane protein. It is still unknown whether GtrIV interacts *in vivo* with other host factors. By utilising the membrane yeast two-hybrid (MYTH) assay, one can study whether GtrIV interacts with GtrA, GtrB and/or other host proteins. MYTH adapts the principle of split ubiquitin for use as a potent *in vivo* sensor of protein–protein interactions, allowing large scale screening for interactors of full-length membrane proteins, from a range of organisms, using *Saccharomyces cerevisiae* as a host (Snider *et al.,* 2010). It offers two major advantages over the conventional yeast two-hybrid system. Firstly, the interaction is detected *in situ* at the cellular membrane instead of the nucleus. Secondly, whole membrane proteins can be assayed. In this system, the bait membrane protein (GtrIV) is fused to the C-terminal half of ubiquitin and an artificial transcription factor. The mutated N-terminal moiety of ubiquitin is fused to the prey protein. Upon interaction of bait and prey proteins, ubiquitin is reconstituted and further recognized by ubiquitin-specific proteases, which subsequently cleave off the transcription factor, thus resulting in reporter gene activation (Snider *et al.,* 2010; Petschnigg *et al.,* 2012)(Figure 7.1). The reporter genes used are two auxotrophic growth markers (HIS3 and ADE2), whose activation enables the yeast to grow on defined minimal medium lacking histidine or adenine, and *lacZ*, encoding the enzyme β - galactosidase for colour development (Snider *et al.,* 2010; Petschnigg *et al.,* 2012). Thus, potential interaction between the two proteins at the membrane of the yeast will be translated into a transcriptional readout, resulting in growth of yeast on selective medium and colour development in a β-galactosidase assay.

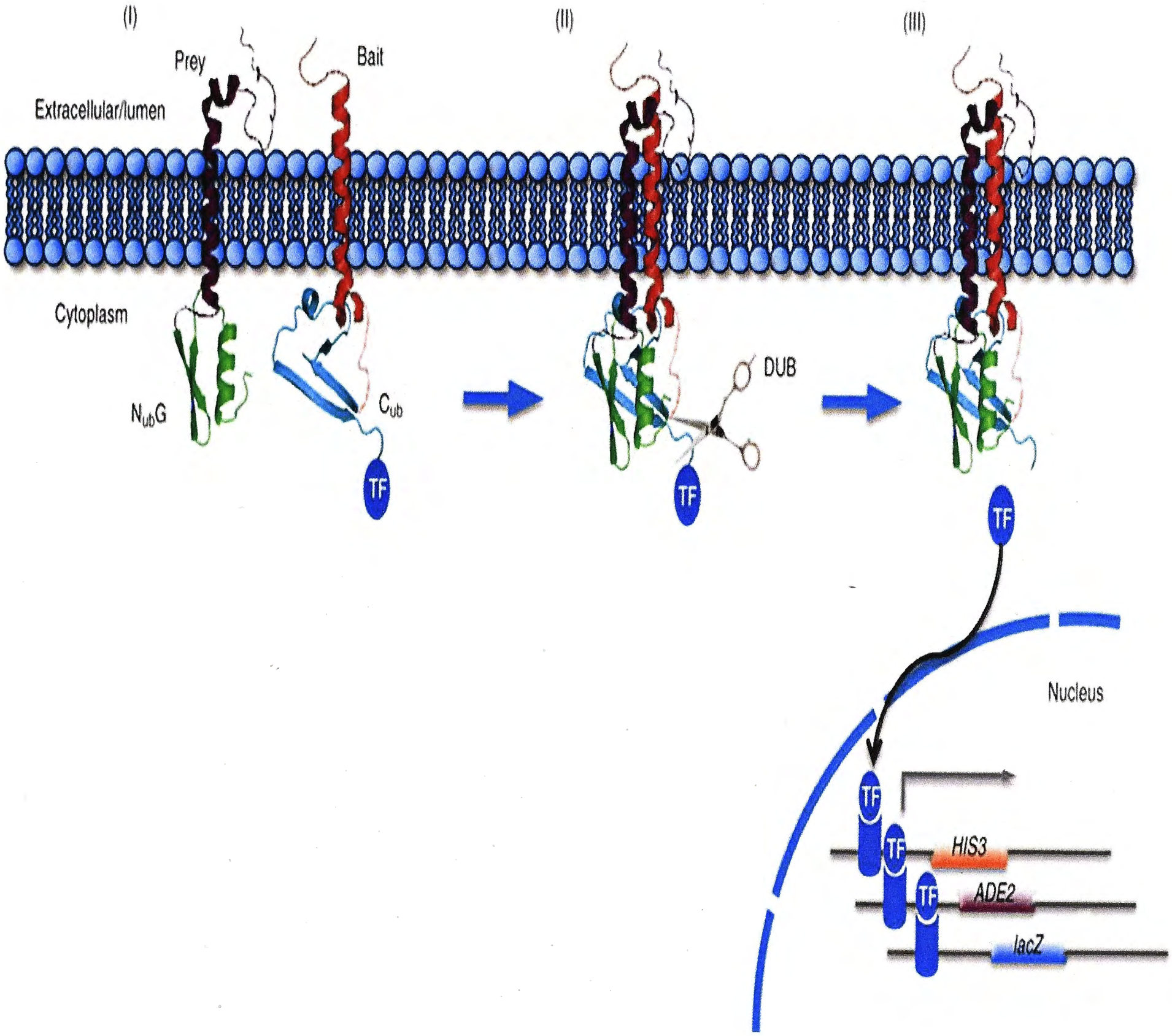

(I)
(II)
(III)
Prey
Bait
Extracellular/lumen
Cytoplasm
$N_{ub}G$
C_{ub}
TF
DUB
Nucleus
HIS3
ADE2
lacZ

Figure 7.1: The membrane yeast two-hybrid assay (MYTH). In the MYTH system, interactions between two membrane proteins are monitored by an indirect transcriptional output. The auxotrophic growth markers are *HIS3*(red) and *ADE2* (purple), whose activation enables the yeast to grow on defined minimal medium lacking histidine or adenine, and *lacZ*, encoding the enzyme β-galactosidase for colour development in the presence of its substrate X-gal. (I) The bait, a membrane protein of interest (shown in red) is fused to Cub followed by the artificial transcription factor LexA-VP16 (TF). The prey, another membrane protein (shown in purple) is fused to the NubG domain. If the bait and the prey membrane proteins which are fused to NubG and Cub respectively do not interact, then the TF is never cleaved from Cub and remains associated with the membrane. This results the genes *HIS3, ADE2* and *lacZ* remaining silent. This renders cells unable to grow on medium lacking histidine and adenine and they are colourless when exposed to X-gal. (II) Interaction between bait and prey proteins allows NubG and Cub to associate and form a pseudoubiquitin molecule, which can then be recognized by cytosolic deubiquitinating enzymes (DUBs). DUB cleavage (scissors) after the C terminus of the Cub tag releases the TF. (III) The TF can then enter the nucleus and activate the reporter system. This turns on the expression of *HIS3*, *ADE2* and *lacZ* resulting in cells which are able to grow on medium lacking histidine and adenine and also turn blue in the presence of X-gal. (Adapted from Snider *et al*, 2010). NubG, N-terminal fragment of ubiquitin with its isoleucine residue replaced to glycine. Cub, C-terminal fragment of ubiquitin.

In order to identify other host factors that may interact with GtrIV (bait), prey libraries can be made by either using *S. flexneri* genomic DNA or generating *S. flexneri* cDNA libraries. For genomic libraries, the *S. flexneri* genomic DNA is randomly cut, size-selected, and the resulting fragments ligated into one or more two-hybrid prey vector(s). Only colonies which display robust growth and a blue colour are selected for further analysis (Rajagopala & Uetz, 2011; Petschnigg *et al.,* 2012). Plasmid DNA can be isolated from these colonies after they have been grown overnight in SD-Tryptophan media only, to select for retention of prey and not bait plasmids (Rajagopala & Uetz, 2011; Petschnigg *et al.,* 2012). The isolated plasmid can then be transformed, minipreped and sequenced using a primer complementary to sequence within the $N_{ub}G$. From this data, a preliminary list of proteins that interact with GtrIV can be generated.

Another way to investigate protein-protein interactions is through the fluorescence resonance energy transfer (FRET) assay. FRET is a physical process in which energy is transferred non-radioactively from an excited fluorophore, serving as a donor, to another chromophore (acceptor) (Masi *et al.,* 2010). Among the techniques related to fluorescence microscopy, FRET is unique in providing signals sensitive to intra- and intermolecular distances in the 1-10 nm range (Masi *et al.,* 2010; Fernández-Dueñas *et al.,* 2012). This assay is based on using fluorescence lifetime imaging (FLIM) to measure FRET between two fluorophores (Llères *et al.*, 2007; Fernández-Dueñas *et al.,* 2012). FRET occurs when the fluorescence emission spectrum of a donor fluorophore overlaps the absorption spectrum of an acceptor fluorophore and their dipoles align (Barber *et al.*, 2009). A schematic of the FRET assay is shown in Figure 7.2. In order to perform this assay, fluorescent protein (FP) tags must be fused to the two proteins believed to interact. In this case, GtrIV and either GtrA or GtrB. These FP tags must be able to function as a FRET pair with one of the pair functioning as a donor and the other as an acceptor. Interaction of the proteins of interest brings the

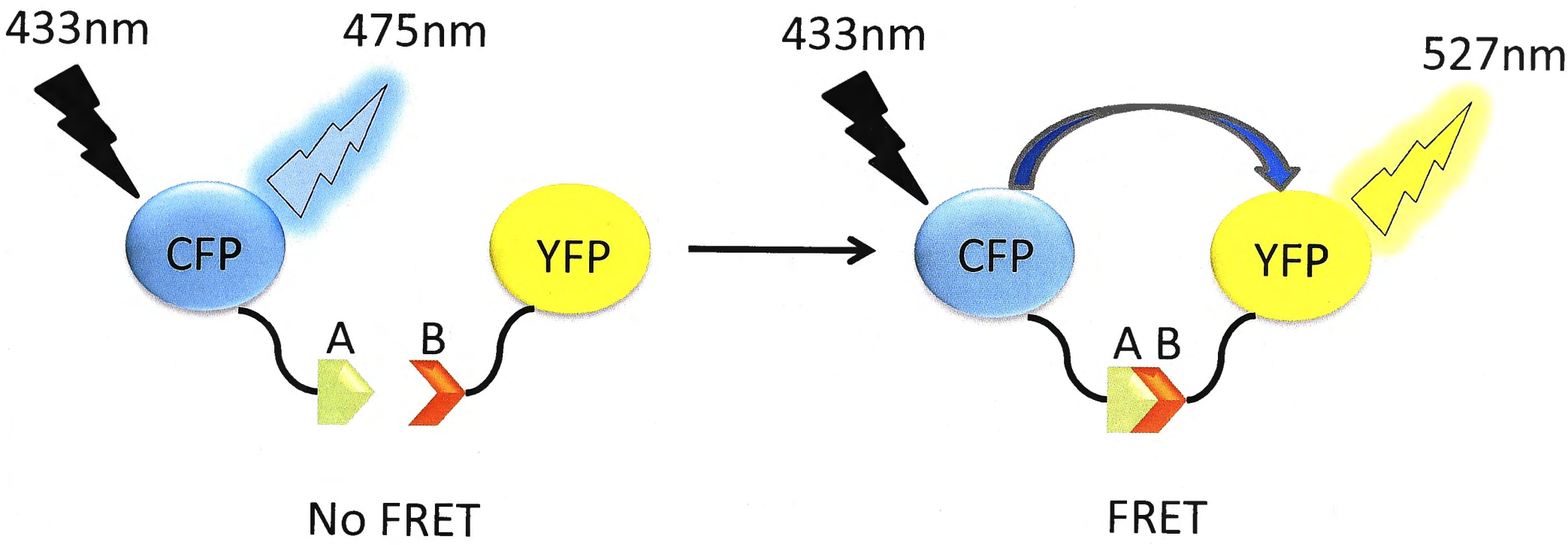


Figure 7.2: Schematic representation of the basic FRET principle applied to the study of plasma membrane proteins. The proteins of interest (**A** and **B**) are fused to the donor and acceptor fluorophores, CFP and YFP, respectively. If A and B do not interact, the donor is excited at 433nm and emits at 475nm. However, if A and B do interact, then the donor and acceptor fluorophores might be in close proximity (⩽10 nm) and energy transfer between the two fluorophores can occur after donor excitation at 433 nm (which emits at 475 nm) and acceptor emission at 527 nm. Adapted from Fernández-Dueñas *et al.,* (2012).

two FP tags into close proximity allowing FRET to occur. FRET can then be detected and measured using FLIM. When performing FRET experiments, care must be also taken to the method chosen for labelling interacting proteins. Two principal tools can be applied: (1) fluorophore tagged antibodies; (2) recombinant fluorescent fusion proteins (Masi *et al.,* 2010; Fernández-Dueñas *et al.,* 2012). The latter method essentially takes advantage of the discovery and use of spontaneously fluorescent proteins, like the green fluorescent protein (GFP). Because of its potency, FRET has been widely used to analyze the structural characteristics of several proteins, including integrins and ion channels. More recently, this method has been applied to clarify the interaction dynamics of these classes of membrane proteins with cytosolic signalling proteins (Barber *et al.,* 2009; Masi *et al.,* 2010).

Obtaining high resolution structural information is a crucial step in elucidating the molecular basis for the function of GtrIV. To help achieve this, one could employ solid-state NMR, which is an extremely versatile and powerful method for determining membrane proteins structures. It is also capable of characterizing the dynamics of these membrane proteins in lipid environments that closely mimic the cell membrane. The other major advantage of using solid-state NMR is that no crystal is required for structure determination (Baldus, M 2007; Etzkorn *et al.,* 2007). The GtrIV protein has to be overexpressed, labeled and purified. Following this, proteoliposomes containing labeled GtrIV can be prepared according to the protocols required for solid-state NMR experiments as described by Andronesi *et al* (2005). The samples can then be subjected to NMR experiments employing MAS (magic-angle-spinning) NMR. The spectra obtained from the MAS-based NMR can be utilized to construct a structural model of GtrIV protein in lipid bilayers.

Alternatively, high resolution structural information can also be achieved by X-ray crystallography (Yaffe MB, 2005; Geerlof *et al.*, 2006). This technique involves the overexpression and subsequent purification and crystallisation of the protein of interest, in this case GtrIV (Geerlof *et al.*, 2006). Following this, X-ray beams are passed through the purified crystal. Due to the intensity and angles of the diffracted beams, a 3-D electron density map can be generated (Hickman & Davies, 2001). This map will help in identifying the residues and catalytic sites involved in O-antigen modification. This would be greatly beneficial in understanding the function of GtrIV. As the natural abundance of membrane proteins is too low for convenient isolation of enough material for structural and functional studies, optimising the overexpression of GtrIV would be an important first step (Andronesi *et al.,* 2005; Wagner *et al.,* 2006; Wagner *et al.,* 2008). Most membrane protein overexpression leads to the production of more protein than the Sec translocon can process. As the Sec translocon is responsible for mediating the translocation of secretory proteins across and the integration of membrane proteins into the cytoplasmic membrane, its saturation makes it impossible for most overexpressed membrane proteins to insert into the membrane (Wagner *et al.,* 2008). Membrane proteins that cannot insert into the membrane end up in the cytoplasm and may aggregate, resulting in hampered translocation of secretory proteins. This adds further stress onto the cell (Wagner et al., 2006; Wagner *et al.,* 2008). In the commonly used *E. coli* strain for expression studies, BL21(DE3), the bacteriophage T7 RNA polymerase (T7RNAP) usually drives recombinant protein production and is controlled by the strong promoter *lac*UV5 (Andronesi *et al.,* 2005). Wagner *et al.,* (2006) reported that key mutations in this promoter led to improved protein expression. This discovery led to Wagner *et al.,* (2008) engineering a derivative strain of *E. coli* BL21(DE3), termed Lemo21(DE3), in which the activity of the T7 RNA polymerase can be precisely controlled by its natural inhibitor T7 lysozyme (T7Lys). By placing T7Lys under the control of an L-rhamnose-inducible marker, the authors showed that adding different amounts of L-

rhamnose into the cultures resulted in different and scalable concentrations of overexpressed protein YidC. As such, this strain may prove useful in overexpressing large quantities of GtrIV for X-ray crystallography.

7.6 Vaccine development

One of the major considerations in using cocktails of live attenuated vaccine strains is to be able to establish O-modification resulting in all possible serotypes. It has been mooted that the creation of a single plasmid that carries the genes of all the Gtrs allowing their expression from the chromosome was a possible solution to the creation of such a multi-valent vaccine. One major drawback in this approach is the potential stability of all the Gtrs expressed in the inner membrane and their competition with one another. Based on the site-directed mutagenesis results for GtrIV loop No. 2, failure to find any critical acidic residues has hinted that loop No. 2 of GtrIV may not play the same conserved role in interacting with the UDP-Glucose as is hypothesised with the other Gtrs. These results coupled by the failure of the GtrIV-GtrIc loop swap chimeras suggests to us that creating chimeric proteins for vaccine work may not be the best way forward at this time. Instead, the best way forward would be to obtain the 3-D structure of GtrIV first. This will allow us to better understand the protein. Then, by identifying the important regions of the protein through its 3-D model one can create functioning chimeric proteins with the other Gtrs, enabling us to create a cocktail of chimeric proteins that express all the various serotypes by ensuring that each chimera consists of two or more closely related Gtrs (GtrI & GtrII, GtrV&GtrX).

7.7 Conclusion

This study has determined the topology of GtrIV to contain 8 transmembrane helices, a large N-terminal periplasmic loop No. 2, a large C-terminal periplasmic loop No. 6, a short cytoplasmic

C-terminal end and a re-entrant loop located after transmembrane helix 3. Through a combination of loop deletion studies and site-directed mutagenesis, the acidic residues D261, E262 and D267 (located in loop No. 6) were critical for GtrIV function when knocked out together. Membrane protein western blotting has revealed that this triple mutant, although non-functional, is able to assemble in the plasma membrane in quantities comparable to that of the unmodified GtrIV. The failure to find any critical acidic residues within GtrIV loop No. 2 indicates that there is a possibility that the acidic residues residing there may also work in combination of one another. Loop swap chimeras were created between GtrIV and its closest structural homologue, GtrIc, to determine if loops No. 2 of both Gtrs play a similar role in glucosylation while their C-terminal periplasmic loops are involved in the conferring serotype specificity. The resulting hybrids lost their native function and were unable to substitute function to the other protein. This signifies the importance of both loops in GtrIV function.

A total of 28 wildtype strains isolated from Japan, Bangladesh and Vietnam that were typed as serotype 4a, 4b or just 4 were subjected to a series of slide agglutination experiments to confirm their serotype. A number of strains had atypical agglutination habits and thus highlighted the possibility that they are potentially new serotypes or subserotypes. Sequencing results revealed the presence of point mutations within several strains, although we were not able to identify mutations that were detrimental to the function of GtrIV. Southern blotting revealed the possibility of two different serotype 4a strains (one of them being possibly derived from a serotype 4b background) circulating within Bangladesh and Japan. Further investigations based on the data provided from this study can lead to better understanding on how $Gtr_{(type)}$ proteins modify the O-antigen in *S. flexneri*, thus advancing our knowledge on this prevalent human pathogen and leads to the development of a multi-valent vaccine against multiple *S. flexneri* serotypes.

Bibliography

Adams, M. M., Allison, G. E. and Verma, N. K. (2001). Type IV O antigen modification genes in the genome of *Shigella flexneri* NCTC 8296. *Microbiology* **147**: 851-860.

Aderem, A. and Ulevitch, R.J. (2000). Toll-like receptors in the induction of the innate immune response. **406**: 782-787.

Adhikari, P., Allison, G., Whittle, B. and Verma, N.K. (1999). Serotype 1a O-antigen modification: Molecular characterisation of the genes involved and their novel organisation in the *Shigella flexneri* chromosome. *Journal of Bacteriology.* **181**:4711-4718.

Ahmed, S. F., Riddle, M. S., Wierzba, T. F., Messih, I. A., Monteville, M. R., Sanders, J. W. and Klena, J. D. (2006) Epidemiology and genetic characterization of *Shigella flexneri* strains isolated from three paediatric populations in Egypt (2000-2004).Epidemiology and Infection. **134:** 1237-1248.

Allison, G. E. and Verma, N. K. (2000) Serotype-converting bacteriophages and O-antigen modification in Shigella flexneri. *Trends in Microbiology*. **8**: 17-23.

Allison, G. E., Angeles, D., Tran-Dinh, N. and Verma, N. K. (2002). Complete genomic sequence of SfV, a serotype-converting temperate bacteriophage of Shigella flexneri. *Journal of Bacteriology*. **184**: 1974-1987.

Al Jarousha, A. M. K., El Jarou M. A. and El Qouqa, I. A., (2010). Bacterial Enteropathogens and Risk Factors Associated with Childhood Diarrhea. *Indian Journal of Pediatrics*: 1-6.

Alexeyev, M. F. and Winkler, H. H. (1999). Membrane topology of the Rickettsia prowazekii ATP/ADP translocase revealed by novel dual pho-lac reporters. *Journal of Molecular Biology*: 1503-1513.

Alexeyev, M. F. and Winkler, H. H. (2002). Transposable dual reporters for studying the structure-function relationships in membrane proteins: permissive sites in *R. prowazekii* ATP/ADP translocase. *Biochemistry*. **41**: 406-414.

Andronesi, O. C., Becker, S., Seidel, K., Heise, H., Young, H. S. and Baldus, M. (2005). Determination of membrane protein structure and dynamics by magic-angle-spinning solid-state NMR spectroscopy. *Journal of the American Chemical Society*. **127**: 12965-12974.

Ashida, H., Ogawa, M., Kim, M., Suzuki, S., Sanada, T., Punginelli, C., Mimuro, H,. and Sasakawa, C. (2011). Shigella deploy multiple countermeasures against host innate immune responses. *Current Opinion in Microbiology* **14**: 16-23.

Ashkenazi, S., Levy, I., Kazaronosvski, V. and Samra, Z. (2003) Growing antimicrobial resistance of Shigella isolates. *Journal of Antimicrobial Chemotherapy* **51**: 427-429.

Badia, J., Ibanez, E., Sabate, M., Baldoma, L., Aguilar, J. (1998). A Rare 920-Kilobase Chromosomal Inversion Mediated by IS1 Transposition Causes Constitutive Expression of the yiaK-S Operon for Carbohydrate Utilization in Escherichia coli. *Journal of Biological Chemistry*. **273**. 8376-8381.

Baldus, M. (2007). The EBSA prize lecture: Magnetic Resonance in the solid state: Applications to protein folding, amyloid proteins and membrane proteins. *European Biophysics Journal*. **36**: 37-48.

Barber, P. R., Ameer-Beg, S. M., Gilbey, J., Carlin, L. M., Keppler, M., Ng, T. C. and Vojnovic, B. (2009). Multiphoton time-domain fluorescence lifetime imaging microscopy: Practical application to protein-protein interactions using global analysis. *Journal of the Royal Society Interface*. **6**: S93-S105.

Barnoy, S., Jeong , K. I., Helm, R. F., Suvarnapunya, A. E., Ranallo, R. T., Tzipori S. and Venkatesan, M. M. (2010). Characterization of WRSs2 and WRSs3, new second-generation virG(icsA)-based Shigella sonnei vaccine candidates with the potential for reduced reactogenicity. *Vaccine* **28**: 1642-1654.

Barreiro, L. B., Laval, G., Quach, H., Patin, E., and Quintana-Murci, L. (2008). Natural selection has driven population differentiation in modern humans. *Nature Genetics*. **40**: 340-345.

Bastin, D. A., Lord, A, and Verma, N. K. (1997). Cloning and analysis of the glucosyl transferase gene encoding type I antigen in *Shigella flexneri*. *FEMS Microbiology Letters*. **156:** 133-139.

Baulard, A. R., Gurcha, S. S., Engohang-Ndong, J., Gouffi, K., Locht, C., Besra, G. S. (2003) In vivo interaction between the polyprenol phosphate mannose synthase Ppm1 and the integral membrane protein Ppm2 from Mycobacterium smegmatis revealed by a bacterial two-hybrid system. *Journal of Biological Chemistry*. **278**: 2242-2248.

Beatty, W. L. and Sansonetti, P. J., (1997). Role of lipopolysaccharide in signalling to subepithelial polymorphonuclear leukocytes. Infection and Immunity **65**: 4395-4404.

Bennish, M. L., (1991). Potentially lethal complications of shigellosis. *Reviews of Infectious Diseases* **13**: S319-S324.

Biet, F., Locht, C. and Kremer, L. (2002). Immunoregulatory functions of interleukin 18 and its role in defense against bacterial pathogens. *Journal of Molecular Medicine*. **80**:147-162.

Birnboim, H. C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*. **7**: 1513-1523.

Bogdanov, M., Zhang, W., Xie, J. and Dowhan, W. (2005) Transmembrane protein topology mapping by the substituted cysteine accessibility method (SCAM™): Application to lipid-specific membrane protein topogenesis. *Methods*. **36**: 148-171.

Bopp, C. A., Brenner, F. W., Fields, P. I., Wells J. G. and N. A. Strockbine. (2003). Escherichia, Shigella, and Salmonella. *Manual of clinical microbiology*. **1**: 654-671.

Bourne, Y., Henrissat, B. (2001). Glycoside hydrolases and glycosyltransferases: families and functional modules. *Current Opinons in Structural Biology*. **11**: 593-600.

Breton, C. and Imberty, A. (1999). Structure/function studies of glycosyltransferases. *Currrent Opinions in Structural Biology*. **9**: 563-571.

Brocke, L., Bendahan, A., Grunewald, M., Kanner, B. I. (2002). Proximity of two oppositely oriented reentrant loops in the glutamate transporter GLT-1 identified by paired cysteine mutagenesis. *Journal of Biological Chemistry*. **277**: 3985-3992.

Brockhausen, I., Hu, B., Liu, B., Lau, K., Szarek, W., Wang, L. and Feng, L. (2008). Characterization of Two 1,3-Glucosyltransferases from Escherichia coli Serotypes O56 and O152. *Journal of Bacteriology*. **190**: 4922-4932.

Brockhausen, I., Hu, B., Liu, B., Lau, K., Szarek, W., Wang, L. and Feng, L. (2008). Characterization of Two 1,3-Glucosyltransferases from Escherichia coli Serotypes O56 and O152. *Journal of Bacteriology*. 4922-4932.

Buchrieser, C., Glaser,PRusniok, C., Nedjari, H., d'Hauteville, H., Kunst, F., Sansonetti, P. and Parsot, C. *(2000)*. The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of Shigella flexneri. *Molecular Microbiology*. ***38:****760-*771.

Buck, K.J. and Amara, S.G. (1994). Chimeric dopamine-norepinephrine transporters delineate structural domains influencing selectivity for catecholamines and 1-methyl-4-phenylpyridinium. *Proceedings of the National Academy of Sciences*. **91**:12584-12588.

Butler, T. (2012). Haemolytic uraemic syndrome during shigellosis. *Trasactions of the Royal Soceity of Tropical Medicine and Hygiene*. **106**: 395-399.

Carlin, N. I., Rahman, M., Sack, D. A., Zaman, A., Kay, B. and Lindberg, A. A. (1989). Use of monoclonal antibodies to type Shigella flexneri in Bangladesh. *Journal of Clinical Microbiology*. **27**:1163–1166.

Carter, J. A., Blondel, C. J., Zaldivar, M., Ålvarez, S. A., Marolda, C. L., Valvano, M. A.. and Contreras, I. (2007). O-antigen modal chain length in Shigella flexneri 2a is growth-regulated through RfaH-mediated transcriptional control of the wzy gene. *Microbiology* **153**: 3499-3507.

Casjens, S., Winn-Stapley, D. A., Gilcrease, E. B., Morona, R., Kuhlewein, C., Chua, J. E., Manning, P. A., Inwood, W. and Clark, A. J. (2004). The chromosome of *Shigella flexneri* bacteriophage Sf6: complete nucleotide sequence, genetic mosaicism, and DNA packaging. *Journal of Molecular Biology*. **339**: 379-394.

Charnock, S. J. and Davies, G. J. (1999). Structure of the Nucleotide-Diphospho-Sugar Transferase, SpsA from *Bacillus subtilis*, in Native and Nucleotide-Complexed Forms. *Biochemistry*. **38**:6380–85

Cheasty, T. and Rowe, B. (1983). Antigenic relationships between the enteroinvasive Escherichia coli O antigens O28ac, O112ac, O124, O136, O143, O144, O152, and O164 and Shigella O antigens. *Journal of Clinical Microbiology*. **17**: 681-684.

Chen Y., Smith, M. R., Thirumalai, K. and Zychlinsky, A. (1996). A bacterial invasin induces macrophage apoptosis by binding directly to ICE. *EMBO Journal*. **15**:3853-3860.

Chen, J. H., Hsu, W. B., Chiou, C. S. and Chen, C. M. (2003). Conversion of *Shigella flexneri* serotype 2a to serotype Y in a shigellosis patient due to a single amino acid substitution in the protein product of the bacterial glucosyltransferase gtrII gene. *FEMS Microbiology Letters*. **224** 277-283.

Christopher, P. R., David, K. V., John, S. M. and Sankarapandian ,V. (2010). Antibiotic therapy for Shigella dysentery. *The Cochrane Library* **8**: 1-104.

Clark, C. A., Beltrame, J. and Manning, P. A. (1991). The oac gene encoding a lipopolysaccharide O-antigen acetylase maps adjacent to the integrase-encoding gene on the genome of *Shigella flexneri* bacteriophage Sf6. Gene. **107**: 43-52.

Clemens, J.D., Stanton, B., Stoll, B., Shahid, N.S., Banu ,H. and Chowdhury, A.K.M.A. (1986). Breast feeding as a determinant of severity in shigellosis. *American Journal of Epidemiology*. **123**:710-720.

Cole, S. T., Supply, P. and Honoré, N. (2001). Repetitive sequences in Mycobacterium leprae and their impact on genome plasticity. *Leprae Review*. **72**: 449-461.

Cosgriff, A. J., Brasier, G., Pi, J., Dogovski, C., Sarsero, J.P,, Pittard, A. J. (2000) A study of AroP-PheP chimeric proteins and identification of a residue involved in tryptophan transport. *Journal of Bacteriology*. **182**: 2207-2217.

Cserzo, M. E., Wallin, I., Simon, G., von Heijne, G. and Elofsson, A. (1997). Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Engineering Design and Selection*. **10**: 673-676.

Daneman, R. and Rescigno, M. (2009). The gut immune barrier and the blood-brain barrier: are they so different? *Immunity*. **31**: 722-735.

Day, W. A., Jr., Fernandez, R. E. and Maurelli, A. T. (2001). Pathoadaptive mutations that enhance virulence: genetic organization of the *cadA* regions of *Shigella* spp. *Infection and Immunity*. **69:** 7471-7480.

DeFelice, L.J., Blakely, R.D. (1996). Pore models for transporters?. *Biophyical Journal*. **70**: 579-80.

Dean, P., (2011) Functional domains and motifs of bacterial type III effector proteins and their roles in infection. *FEMS Microbiolical Review*.

Deane, J. E., Roversi, P., Cordes, F. S., Johnson, S., Kenjale, R., Daniell, S., Booy, F., Picking, W.D., Picking, P. L., Blocker, A. J. and Lea, S. M. (2006). Molecular model of a type III secretion system needle: Implications for host-cell sensing. *PNAS*. **103**: 12529-12533.

Deane, J. E., Graham, S. C., Mitchell, E. P., Flot, D., Johnson, S. and Lea, S. M. (2008). Crystal structure of Spa40, the specificity switch for the Shigella flexneri type III secretion system. *Molecular Microbiology*. **69**: 267-276.

Dickenson, N. E., Zhang, L., Epler, C. R., Adam, P. R., Picking, W. L. and Picking, W. D. (2011) Conformational changes in IpaD from shigella flexneri upon binding bile salts provide insight into the second step of type III secretion. *Biochemistry*. **50**: 172-180.

Dower, W.J., Miller, J. F. and Ragsdale, C. W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Research*. **16:**6127-6145.

Doyle, M. P. and Padhye, V. V. *(*1989). Escherichia coli. *Foodborne bacterial pathogens*. 235-281.

Doyle, D. A., Morais Cabral, J., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T. and MacKinnon, R. (1998). The structure of the potassium channel: molecular basis of K+ conduction and selectivity. *Science*. **280**: 69-77.

Driessen, A. J. M., Manting, E. H. and Van Der Does, C. (2001). The structural basis of protein targeting and translocation in bacteria. *Nature Structural Biology* **8**: 492-498.

DuPont, H., Levine, M. M., Hornick, R. and Formal, S. (1989). Inoculum size in shigellosis and implications for expected mode of transmission. *Journal of Infectious Diseases* **159**: 1126-1128.

Dyer, M. D., Neff, C., Dufford, M., Rivera, C. G., Shattuck, D., Bassaganya-Riera, J., Murali, T. M. and Sobral, B. W. (2010). Analysis of protein-protein interactions using high-throughput yeast two-hybrid screens. *PLoS One*. **5**: e12089.

El-Gendy, A., El-Ghorab, N., Lane, E. M., Elyazseed, R. A., Carlin, N. I. A., Mitry, M. M., Kay, B. A., Savarino, S. J. and Peruski, L. F. Jnr. (1999). Identification of Shigella flexneri subserotype 1c in rural Egypt. *Journal of Clinical Microbiology*. **37:**873–874.

Erridge, C., Bennett-Guerrero E. and Poxton, I. R. (2002). Structure and function of lipopolysaccharides. *Microbes and Infection* **4**: 837-851.

Escobar-Paramo, P., Giudicelli, C., Parsot, C. and Denamur, E. (2003). The evolutionary history of *Shigella* and enteroinvasive *Escherichia coli* revised. Journal of Molecular Evolution. **57:** 140-148

Espina, M., Ausar, S. F.,Middaugh, C. R., Pickering, W. D. and Pickering, W. L. (2006). Spectroscopic and calorimetric analyses of invasion plasmid antigen D (IpaD) from *Shigella flexneri* reveal the presence of two structural domains.*Biochemistry*. **45**: 9219-27.

Etzkorn, M., Martell, S. andronesi, O. C., Seidel, K., Engelhard, M. and Baldius, M. (2006). Secondary structure, dynamics, and topology of a seven-helix receptor in native membranes, studied by solid-state NMR spectroscopy. *Angewandte Chemie International Edition English*. **46**: 459-462.

Farshad, F., Sheikhi, R., Janponi, A., Jaziri, E. and Alborsi, A. (2006). Characterization of *Shigella* strains in Iran by plasmid profile analysis and PCR amplification of ipa genes. *Journal of Clinical Microbiology*. **8**:2879-83.

Faherty, C. S. and Maurelli, A.T. (2009) Spa15 of *Shigella flexneri* Is Secreted through the Type III Secretion System and Prevents Staurosporine-Induced Apoptosis. *Infection and Immunity*. **77**: 5281-5290.

Fasano, A., Norieha, F. R., Maneal Jr, D. R., Chanasongcram, S., Russell, R., Guandalini S. and Levine, M. M. (1995). *Shigella* enterotoxin 1: an enterotoxin of *Shigella flexneri* 2a active in rabbit small intestine in vivo and in vitro. *Journal of Clinical Investigation*. **95:** 2853–2861.

Fasano, A., Noriega, F. R., Liao, F. M., Wang, W. and Levine, M. M. (1997). Effect of *Shigella* enterotoxin 1 (ShET1) on rabbit intestine in vitro and in vivo. *Gut*. **40**: 505–511.

Fekkes, P. and Driessen, A. J. M. (1999). Protein targeting to the bacterial cytoplasmic membrane. *Microbiology and Molecular Biology Reviews* **63**: 161-173.

Fernández-Dueñas, V., Llorente, J., Gandía, J., Borroto-Escuela, D. O., Agnati, L.F., Tasca, C. I., Fuxe, K., Ciruela, F. (2011). Fluorescence resonance energy transfer-based technologies in the study of protein–protein interactions at the cell surface. *Methods Molecular Biology*. **781**: 1-29.

Ferreccio, C., Prado, V., Ojeda, A., Cayyazo, M., Abrego, P., Guers, L. and Levine M. M. (1991). Epidemiologic patterns of acute diarrhea and endemic Shigella infections in children in a poor periurban setting in Santiago, Chile. *American Journal of Epidemiology* **134**: 614-627.

Fetchko, M. and I, Stagljar. (2004). Application of the split-ubiquitin membrane yeast two-hybrid system to investigate membrane protein interactions. *Methods*. **32**: 349-362.

Fiedler, S., Broecker, J. and Keller, S. (2010) Protein folding in membranes. *Cellular and Molecular Life Sciences: CMLS*. **67**: 1779-1798.

Fukushima, M., Kakinuma, K. and Kawaguchi, R. (2002). Phylogenetic analysis of Salmonella, Shigella, and Escherichia coli strains on the basis of the gyrB gene sequence. Journal of Clinical Microbiology. **40:** 2779-2785.

Gala´ n, J. E. and Wolf-Watz, H. (2006). Protein delivery into eukaryotic cells by type III secretion machines. *Nature*. **444**: 567.

Garinot-Schneider, C., Lellouch, A. C. and Geremia, R. A. (2000). Identification of Essential Amino Acid Residues in *Sinorhizobium meliloti* Glucosyltransferase ExoM. *Journal of Biological Chemistry*. **275**: 31407-31413.

Geerlof, A., Brown, J., Coutard, B., Egloff, M. P., Enguita, F. J., Fogg, M. J., Gilbert, R. J. C., Groves, M. R., Haouz, A., Nettleship, J. E., Nordlund, P., Owens, R. J., Ruff, M., Sainsbury, S., Svergun, D. I. and Wilmanns, M. (2006). The impact of protein characterization in structural proteomics. *Acta Crystallographica Section D: Biological Crystallography*. **62**: 1125-1136.

Girardin, S. E., Boneca, I. G. Viala, J., Chamaillard, M., Labigne, A., Thomas, G., Philpot, D. J. and Sansonetti, P. J. (2003). Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. *Journal of Biological Chemsitry*. **14**: 8869-8872.

Giron, J. A. (1995). Expression of flagella and motility in *Shigella*. *Molecular Microbiology*. **18**: 63-75

Giros, B., Wang, Y. M., Suter, S., McLeskey, S. B., Pifl, C., Caron, M. G. (1994). Delineation of discrete domains for substrate, cocaine, and tricyclic antidepressant interactions using chimeric dopamine-norepinephrine transporters. *Journal of Biological Chemistry*. **269**:15985-15988.

Grunewald, M., Bendahan, A. and Kanner, B. I. (1998). Biotinylation of single cysteine mutants of the glutamate transporter GLT-1 from rat brain reveals its unusual topology. *Neuron*. **21**: 623-632.

Grunewald, M., Menaker, D. and Kanner, B. I. (2002): Cysteine-scanning mutagenesis reveals a conformationally sensitive reentrant pore-loop in the glutamate transporter GLT-1. *Journal of Biological Chemistry*. **277**: 26074-26080.

Guan, S. and Verma, N. K. (1998). Serotype conversion of a *Shigella flexneri* candidate vaccine strain via a novel site-specific chromosome-integration system. *FEMS Microbiology Letters*. **166**. 79-87.

Guan, S., Bastin, D. A. and Verma, N. K. (1999). Functional analysis of the O antigen glucosylation gene cluster of Shigella flexneri bacteriophage SfX. *Microbiology* **145**: 1263-1273.

Hale, T. L. (1991). Genetic basis of virulence in Shigella species. *Microbiology Reviews*. **55**: 206-224.

Harrington, S. E. and Ben-Tal, N. (2009). Structural Determinants of Transmembrane Helical Proteins. *Structure* **17**: 1092-1103.

Harris-Warrick, R. M. (2000). Ion channels and receptors: molecular targets for behavioral evolution. *Journal of Computational Physiology A*. **186**: 605-616.

Hickman, A. B. and Davies, D. R. (2001). Principles of macromolecular X-ray crystallography. *Current protocols in protein* science. **Chapter 17**.

Hitchcock, P. J. and Brown, T. M. (1983). Morphological heterogeneity among Salmonella lipopolysaccharide chemotypes in silver-stained polyacrylamide gels. *Journal of Bacteriology*. **154**: 269-277

Hirokawa, T., Boon-Chieng, S. and Mitaku, S. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics. **14**:378-379.

Hofmann, K. and Stoffel, W. (1993). TMbase - A database of membrane spanning protein segments. *The Journal Biological Chemistry*. **374**:166.

Hong, M. and Payne, S.M. (1997). Effect of mutations in Shigella flexneri chromosomal and plasmid-encoded lipopolysaccharaide genes on invasion and serum resistance. *Molecular Microbiology*. **24**: 779-791.

Hong, S., Choi, Y.H., Choo, Y. A., Choi, Y., Choi, S. Y., Kim, D. W., Lee, B. K. and Park, M.S. (2010). Genetic characterization of atypical Shigella flexneri isolated in Korea. *Journal of Microbiology and Biotechnology*. **20**: 1457-1462.

Huan, P. T., Whittle, B. L., Bastin, D. A., Lindberg, A. A. and Verma, N. K. (1997). *Shigella flexneri* type-specific antigen V: cloning, sequencing and characterization of the glucosyl transferase gene of temperate bacteriophage SfV. Gene. **195**: 207-216.

Iijima, H., Takahashi, I. and Kiyono, H. (2001). Mucosal immune network in the gut for the control of infectious diseases. *Reviews in Medical Virology*. **11**:117-33.

Ingersoll, M., Groisman, E. A. and Zychlinsky, A. (2002). Pathogenicity islands of *Shigella*. *Current Topics in Microbiolgy and Immunology*. **264:**49–65.

Ingersoll, M. A. and Zychilinsky, A. (2006). ShiA abrogates the innate T-cell response to *Shigella flexneri* infection. *Infection and Immunity*. **74:** 2317–2327.

Islam, D., Wretlind, B., Ryd, M., Lindberg, A.A., andChristensson, B. (1995). Immunoglobulin subclass distribution and dynamics of *Shigella*-specific antibody responses in serum and stool samples in shigellosis. *Infection and Immunity*. **63**:2054-2061.

Jehl, S. P., Doling, A. M., Giddings, S, K., Phalipon, A., Sansonetti, P. J.,. Goldberg, M. and Starnbach, M. N. (2011) Antigen-Specific $CD8^{+}$ T Cells Fail To Respond to Shigella flexneri *Infection and Immunity.* **79**: 2021-2030.

Jennison, A. V. and Verman, N. K. (2004). Shigella flexneri infection: Pathogenesis and vaccine development. *FEMS Microbiology Reviews* **28**: 43-58.

Jennison, A. V. (2006). Construction of a multivalent vaccine strain of Shigella flexneri and evaluation of serotype-specific immunity. *Immunology Medical Microbiology*. **46**: 444-451.

Jepson, M. A. and Clark, M. A. (1998). Studying M cells and their role in infection. *Trends in Microbiology* **6**: 359-365.

Jin, Q., Yuan, Z., Xu, J., Wang, Y., Shen, Y., Lu, W., Wang, J., Liu, H., Yang, J., Yang, F., Zhang, X., Zhang, J., Yang, G., Wu, H., Qu, D., Dong, J., Sun, L., Xue, Y., Zhao, A., Gao, Y., Zhu, J., Kan, B., Ding, K., Chen, S., Cheng, H., Yao, Z., He, B., Chen, R., Ma, D., Qiang, B., Wen, Y., Hou, Y. and Yu, J. (2002). Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. *Nucleic Acids Research*. **30**: 4432-4441.

Johansson, A. and Lindahl, E. (2006). Amino-Acid Solvation Structure in Transmembrane Helices from Molecular Dynamics Simulations. *Biophysical Journal*. **91**: 4450-4463.

Kaminski, R. W. and Oaks, E. V. (2009). Inactivated and subunit vaccines to prevent shigellosis. *Expert Review of Vaccines*. **8**: 1693-1704.

Kaminski, L. and Eichler, J. (2010). Identification of residues important for the activity of Haloferax volcanii AglD, a component of the archaeal N-glycosylation pathway. *Archaea*. 2010: 1-9.

Kenne, L., Lindberg, B. and Petersson, K. (1977). Basic structure of the oligosaccharide repeating-unit of the *Shigella flexneri* O-antigens. Carbohydrate Research. **56:** 363-370.

Kenne, L., Lindberg, B. and Petersson, K. Katzenellenbogen, E. and Romanowska, E. (1978). Stuctural Studies of *Shigella flexneri* O-Antigens. *European Journal of Biochemistry*. **91**: 279-284.

Kenjale, R.,Wilson, J., Zenk, S. F., Saurya, S., Picking, W. L., Picking, W. D. and Blocker, A. (2005). The Needle Component of the Type III Secreton of *Shigella* Regulates the Activity of the Secretion Apparatus. Journal of Biological Chemistry. **280**: 42929-42937.

Kohler, H., Rodrigues, S.P. and McCormick, B.A. (2002). Shigella flexneri Interactions with the Basolateral Membrane Domain of Polarized Model Intestinal Epithelium: Role of Lipopolysaccharide in Cell Invasion and in Activation of the Mitogen-Activated Protein Kinase ERK. *Infection and Immunity* **70**:1150-1158.

Korres, H. and Verma, N. K. (2004). Topological analysis of glucosyltransferase GtrV of Shigella flexneri by a dual reporter system and identification of a unique reentrant loop. *Journal of Biological Chemistry* **279**: 22469-22476.

Korres, H. (2006). Structural and Functional Analysis of Glucosyltransferase V (GtrV) of *Shigella flexneri* (Honours Thesis).

Korres, H. and Verma, N. K. (2006). Identification of essential loops and residues of glucosyltransferase V (GtrV) of Shigella flexneri. *Molecular Membrane Biology* **23**: 407-419.

Kosek, M., Yori, P. P. and Olortegui, M. P. (2010). Shigellosis update: advancing antibiotic resistance, investment empowered vaccine development, and green bananas. *Current Opinion in Infectious Diseases* **23**: 475-480.

Kotloff, K. L., Winickoff, J. P., Ivanoff, B., Clemens, J. D., Swerdlow, D. L., Sansonetti P. J., Adak, G. K. and Levine, M. M. (1999). Global burden of Shigella infections: Implications for vaccine development and implementation of control strategies. *Bulletin of the World Health Organization* **77**: 651-666.

Kotloff, K.L., Noriega, F.R., Samandari, T., Sztein, M.B., Losonsky, G.A,, Nataro, J.P., Picking, W.D., Barry, E.M., Levine, M.M., Sansonetti, P.J. (2000). Shigella flexneri 2a strain CVD 1207, with specific deletions in virG, sen, set, and guaBA, is highly attenuated in humans. *Infection and Immunity*. **68**: 1034.

Kwaw, I., Sun, J. and Kaback, H.R. (2000) Thiol cross-linking of cytoplasmic loops in the lactose permease of *Escherichia coli. Biochemistry*. **39**: 3134-3140.

Kweon, M. N. (2008) Shigellosis: The current status of vaccine development. *Current Opinion in Infectious Diseases* **21**: 313-318.

Lafont, F., Van Nhieu, G. T., Hanada K., Sansonetti, P. and Van der Goot, F. G., (2002). Initial steps of Shigella infection depend on the cholesterol/sphingolipid raft-mediated CD44-IpaB interaction. *EMBO Journal*. **21**: 4449-4457.

Lan, R. and P. R. Reeves. (2002). Escherichia coli in disguise: molecular origins of Shigella. *Microbes and Infection*. **4:**1125-*1132.*

Lan, R., Alles, M. C., Donohoe, K., Martinez, M. B. and Reeves, P. R. (*2004)*. Molecular evolutionary relationships of enteroinvasive Escherichia coli and Shigella spp. *Infection and Immunity*. **72:**5080-508.

Larkin, M. A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2007). *Bioinformatics*. **23**: 2947-2948.

Lehane, A. M., Korres, H. and Verma, N. K. (2005). Bacteriophage-encoded glucosyltransferase GtrII of Shigella flexneri: Membrane topology and identification of critical residues. *Biochemical Journal* **389**: 137-143.

Lehrer, J., Vigeant, K. A., Tatar, L. D., Valvano, M.A. (2007). Functional characterization and membrane topology of Escherichia coli WecA, a sugar-phosphate transferase initiating the biosynthesis of enterobacterial common antigen and O-antigen lipopolysaccharide. *Journal of Bacteriology*. **189**: 2618-2628.

Lerouge, I. and Vanderleyden, J. (2002). O-antigen structural variation: mechanisms and possible roles in animal/plant-microbe interactions. *FEMS Microbiology Reviews*. **26**: 17-47.

Liapakis, G., Simpson, M. M. and Javitch, J. A. (2001). The substituted-cysteine accessibility method (SCAM) to elucidate membrane protein structure. *Current protocols in* neuroscience. **Chapter 4**.

Llères, D., S. Swift. and Lamond, A. I. (2007) Detecting protein-protein interactions in vivo with FRET using multiphoton fluorescence lifetime imaging microscopy (FLIM). *Current protocols in cytometry*. **Chapter 12**.

Lindberg, A. A. and Pal, T. (1993). Strategies for development of potential candidate *Shigella* vaccines. *Vaccine*. **11**: 168-179.

Mahajan-Miklos, S., Tan, M. W., Rahme, L. G. and Ausubel, F. M. (1999*)*. Molecular mechanisms of bacterial virulence elucidated using a Pseudomonas aeruginosa-Caenorhabditis elegans pathogenesis model. *Cell*. **96:** 47-*56*.

Mahoney, F. J., Farley, T. A., Burbank, D. F., Leslie, N. H. and McFarland, L. M. (1993). Evaluation of an intervention program for the control of an outbreak of shigellosis among institutionalized persons. *Journal of Infectious Diseases* **168**: 1177-1180.

Makela, P. H. and Stocker, B. A,. (1969). How genes determine the structure of the *Salmonella* lipopolysaccharide. *Journal of General Microbiology*. **57**: VI.

MacKinnon, R. (1995). Pore loops: an emerging theme in ion channel structure. *Neuron*. **14**: 889-892.

Mallett, C. P., VanDeVerg, L., Collins, H. H. and Hale, T. L. (1993). Evaluation of *Shigella* vaccine safety and efficacy in an intranasally challenged mouse model. *Vaccine*. **11**: 190-196.

Manoil, C., Mekalanos, J. J. and Beckwith, J. (1990). Alkaline phosphatase fusions: sensors of subcellular location. Journal of Bacteriology. **172**: 515-518.

Mantis, N., Prevost, M.C. and Sansonetti, P. (1996). Analysis of epithelial cell stress response during infection by Shigella flexneri. *Infection and Immunity*. **64**: 2474-2482.

Marteyn, B., West, N. P., Browning, D. F., Cole, J. A., Shaw, J. G., Palm, F., Mounier, J., Prevost, M. C., Sansonetti P. and Tang, C. M. (2010) Modulation of Shigella virulence in response to available oxygen in vivo. *Nature* **465**: 355-358.

Martínez-Alonso, M., Toledo-Rubio, V., Noad, R., Unzueta, U., Ferrer-Miralles, N., Roy, P. and Villaverde, A. (2009). Rehosting of bacterial chaperones for high-quality protein production. *Applied and Environmental Microbiology* **75**: 7850-785.

Martinez-Becerra, F. J., Kissmann, J. M., Diaz-McNair, J., Choudhari, S. P., Quick, A. M., Mellado-Sanchez, G., Clements, J. D., Pasetti, M. F. and Picking, W. L. (2012). A broadly protective *Shigella* vaccine based on Type III secretion apparatus proteins. *Infection and Immunity.* **80**: 1222-1231.

Masi, A., Cicchi, R., Pavone, F. S. and Arcangeli, A. (2010). Optical methods in the study of protein-protein interactions. *Advances in Experimental Medicine and Biology* **674:**33-42

Mattoo, S., Lee, Y. M. and Dixon, J. E. (2007). Interactions of bacterial effector proteins with host proteins. *Current Opinions in Immunogyl.* **19**, 392–401.

Maurelli, A. T. (2007). Black holes, antivirulence genes, and gene inactivation in the evolution of bacterial pathogens. *FEMS Microbiolgy Letters*. **267:**1-8.

Mavris, M., Manning, P. A. and Morona, R. (1997). Mechanism of bacteriophage SfII-mediated serotype conversion in *Shigella flexneri*. Molecular Microbiology. **26:** 939-950.

Mavris, M., Sansonetti, P. J. and Parsot, C. (2002). Identification of the cis-acting site involved in activation of promoters regulated by activity of the type III secretion apparatus in Shigella flexneri. *Journal of Bacteriology*. **184**: 6751-6759.

Mavris, M. and Sansonetti, P. (2004). Microbial-gut interactions in health and disease. Epithelial cell responses. *Best Practice and Research Clinical Gastroenterol*. **18**: 373-86.

Menard, R., Sansonetti, P.J., Parsot, C. and Vasselon, T. (1994) Extracellular association and cytoplasmic partitioning of the IpaB and IpaC invasins of *Shigella flexneri*. *Cell*. **79**:515-525.

Menard, R., Sansonetti, P., Parsot, C. (1994). The secretion of the Shigella flexneri Ipa invasins is activated by epithelial cells and controlled by IpaB and IpaD. *Embo Journal.* **13**: 5293-5302.

Miller, J. H. (1992). A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia Coli and Related Bacteria. New York: *Cold Spring Harbor Laboratory Press*.

Morona, R., Macpherson, D. F., Van Den Bosch, L., Carlin, N. I. and Manning, P. A. (1995). Lipopolysaccharide with an altered O-antigen produced in Escherichia coli K-12 harbouring mutated, cloned Shigella flexneri rfb genes. *Molecular* Microbiology. **18**: 209-223.

Mounier, J., Vasselon, T., Hellio, R., Lesourd, M. and Sansonetti, P. J. (1992). *Shigella flexneri* enters human colonic Caco-2 epithelial cells through the basolateral pole. *Infection and Immunity* **60**: 237-248.

Mounier, J., Popoff, M. R., Enninga, J., Frame, M. C., Sansonetti, P. J. and Van Nhieu G. T. (2009). The IpaC carboxyterminal effector domain mediates Src-dependent actin polymerization during Shigella invasion of epithelial cells. *PLoS Pathogens*. **5**.

Nagy, G. and Pal, T. (2008). Lipopolysaccharide: A tool and target in enterobacterial vaccine development. *Biological Chemistry*. **389**: 513-520.

Nair, A. (2006). Structural and functional analysis of glucosyltransferase IV (GtrIV) of *Shigella flexneri* (Honours Thesis).

Nair, A., Korres, H. and Verma, N. K. (2011). Topological chracterisation and identification of critical domains within glucosyltransferase IV (GtrIV) of *Shgella flexneri*. *BMC Biochemistry*. **12**: 67-81.

Nakata, N., Tobe, T., Fukuda, I., Suzuki, T., Komatsu, K., Yoshikawa M. and Sasakawa, C. (1993). The absence of a surface protease, OmpT, determines the intercellular spreading ability of *Shigella*: the relationship between the *ompT* and *kcpA* loci. *Molecular Microbiology*. **9:** 459-468.

Nhieu ,G. T. V., Ben-Zeev, A., Sansonetti, P.J. (1997). Modulation of bacterial entry into epithelial cells by association between vinculin and the *Shigella* IpaA invasin. *EMBO Journal*. **16**:2717-2729.

Nhieu, G. T. and Sansonetti, P. J., (1999). Mechanism of *Shigella* entry into epithelial cells. *Current Opinions in Microbiology*. **2**: 51-55.

Nicholas, H. R. and Hodgkin, J. (2004). Responses to infection and possible recognition strategies in the innate immune system of Caenorhabditis elegans. *Molecular Immunology*. **41:** 479-493.

Nie, H., Yang, F., Zhang, X., Yang, J., Chen, L., Wang, J., Xiong, Z., Peng, J., Sun, L., Dong, J., Xue, Y., Xu. X., Chen, S., Yao, Z., Shen, Y. and Jin, Q. (2006). Complete genome sequence of *Shigella flexneri* 5b and comparison with *Shigella flexneri* 2a. *BMC Genomics*. **7:** 173.

Niyogi, S. K. (2005). Shigellosis. *Journal of Microbiology*. **43**: 133-143.

Nishizawa, K., Shimoda, E., Kasahara, M. (1995). Substrate recognition domain of the Gal2 galactose transporter in yeast Saccharomyces cerevisiae as revealed by chimeric galactose-glucose transporters. *Journal of Biological Chemistry*. **270**: 2423-2426.

Nutten, S., Sansonetti, P., Huet, G., Bourdon-Bisiaux, C., Meresse, B., Colombel, J. F. and Desreumaux, P. (2002). Epithelial inflammation response induced by *Shigella flexneri* depends on mucin gene expression.Microbes and Infection. **11**: 1121-1124.

Oaks, E.V., Wingfield, M.E. and Formal, S.B. (1985). Plaque formation by virulent *Shigella flexneri*. *Infection and Immunity*. **48**: 124-129.

Ochman, H., Lawrence, J. G. and Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature*. **405**: 299–304.

Ogawa, M. and Sasakawa, C. (2006). Intracellular survival of Shigella. *Cellular Microbiology* **8**: 177-184.

Ogawa, M., Handa, Y., Ashida, H., Suzuki, M. and Sasakawa, C. (2008). The versatility of *Shigella* effectors. *Nature Reviews Microbiology*. **6**: 11-16.

Opal, S. M. (2007). The host response to endotoxin, antilipopolysaccharide strategies, and the management of severe sepsis. *International Journal of Medical Microbiology* **297**: 365-377.

Opintan, J. and M. J. Newman, (2007). Distribution of serogroups and serotypes of multiple drug resistant Shigella isolates. *Ghana Medical Journal* **41**: 8-29.

Oriol, R., Martinez-Duncker, I., Chantret, I., Mollicone, R. and Codogno, P. (2002): Common origin and evolution of glycosyltransferases using Dol-P-monosaccharides as donor substrate. *Molecular Biology and Evolution*. **19**: 1451-1463.

Parsot, C., Ageron, E., Penno, C., Mavris, M., Jamoussi, K., d'Hauteville, H., Sansonetti P. and Demers, B. (2005). A secreted anti-activator, OspD1, and its chaperone, Spa15, are involved in the control of transcription by the type III secretion apparatus activity in Shigella flexneri. Molecular Microbiology. **56:**1627-1635.

Peng, J., Zhang, X., Yang, J., Wang, J., Yang, E., Bin, W., Wei, C., Sun, M. and Jin, Q. (2006). The use of comparative genomic hybridization to characterize genome dynamics and diversity among the serotypes of *Shigella*. *BMC Genomics*. **7:**218.

Perdomo, J. J., Gounon, P. and Sansonetti, P. J. (1994). Polymorphonuclear leukocyte transmigration promotes invasion of colonic epithelial monolayer by *Shigella flexneri*. Journal of Clinical Investigation **93**: 633-643.

Petschnigg, J., Wong, V., Snider, J. and Stagljar, I. (2012). Investigation of Membrane Protein Interactions Using the Split-Ubiquitin Membrane Yeast Two-Hybrid System. *Methods in Molecular Biology*. **812**: 225-244.

Phalipon, A., Cardona, A., Kraehenbuhl , J.P., Edelman, L., Sansonetti, P.J. and Corthesy, B. (2002) Secretory component: a new role in secretory IgA-mediated immune exclusion in vivo. *Immunity*. **17**: 107-115.

Phalipon, A. and P. Sansonetti., (2007). Shigella's way of manipulating host intestinal innate and adaptive immune system: a tool box for survival? *Immunology and Cell Biology*: 1-11.

Philpott, D.J., Edgeworth, J. D. and Sansonetti, P.J. (2000). The pathogenesis of Shigella flexneri infection: lessons from in vitro and in vivo studies. *Philos Trans R Soc Lond B Biological Science*. **355**: 575-586.

Pickering, L. K., Evans, D. G., DuPont, H. L., Vollet, J. J., 3rd and Evans, D. J., Jr. (1981). Diarrhea caused by *Shigella*, rotavirus, and *Giardia* in day-care centers: prospective study. *Journal of Pediatrics* **99**: 51-56.

Pupo, G. M., Karaolis, D. K., Lan, R. and . Reeves, P. R. **(1997)**. Evolutionary relationship among pathogenic and nonpathogenic Escherichia coli strains inferred from multilocous enzyme electrophoresis and mdh sequence studies. *Infection and Immunity*. **65:**2685-2692.

Pupo, G.M., Lan, R. and Reeves, .P.R. (2000). Multiple independent origins of Shigella clones of Escherichia coli and convergent evolution of many of their characteristics. *Proceedings of the National Academy of Sciences*.**97**:10567–10572

Qiu, S., Wang, Z., Chen, C., Liu, N., Jia, L., Liu, W., Wang, L., Hao, R., Zhang, L., Wang, Y. and Song, H. (2011). Emergence of a Novel Shigella flexneri Serotype 4s Strain That Evolved from a Serotype X Variant in China. Journal of Clinical Microbiology. **49**: 1148-1150.

Rahman, M. M. and McFadden, G. (2011). Modulation of NF-κB signalling by microbial pathogens. *Nature Reviews Microbiology*. **9**: 291–306.

Ramakrishnan, B. and Qasba, P. K. (2001). Crystal structure of lactose synthase reveals a large conformational change in its catalytic component, the β1,4-galactosyltransferase-I. *Journal of Molecular Biology*. **310**: 205-218.

Ram, P. K., Crump, J. A., Gupta, S. K., Miller, M. A. and Mintz E. D. (2008) Part II. Analysis of data gaps pertaining to Shigella infections in low and medium human development index countries, 1984-2005. *Epidemiology and Infection* **136**: 577-603.

Ramiscal, R., Tang, S.S., Korres, S. and Verma, N. K. (2010). Structural and functional divergence of the newly identified GtrIc from its Gtr family of conserved Shigella flexneri serotype-converting glucosyltransferases. *Molecular Membrane Biology* **1**: 1-13.

Ramos, H. C., Rumbo, M. and Sirard, J. C. (2004). Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. *Trends in Microbiology*. **12**: 509-517.

Rasolofo-Razanamparany, V., Cassel-Beraud, A., Roux, J., Sansonetti, J.P. and Phalipon, A. (2001). Predominance of serotype-specific mucosal antibody response in Shigella flexneri-Infected humans living in an area of endemicity. *Infection and Immunity*, **69**:5230-5234.

Rathman, M., Jouirhi, N., Allaoui, A., Sansonetti, P., Parsot, C. and Nhieu T. V. G. (2000). The development of a FACS-based strategy for the isolation of *Shigella flexneri* mutants that are deficient in intercellular spread. *Molecular Microbiology*. **35**: 974-90.

Reed, W. P. and Albright,E. L. (1974). Serum factors responsible for killing of Shigella. *Immunology* **26**: 205-215.

Riddle, M. S., Kaminski, R. W., Williams, C., Porter, C., Baqar, S., Kordiš, A., Gilliland, T., Lapa, J., Coughlin, M., Soltis, C., Jones, E., Saunders, J., Keiser, P. B., Ranallo, R. T., Gormley, R., Nelson, M., Turbyfill, K. R., Tribble, D. and Oaks, E. V. (2011). Safety and immunogenicity of an intranasal *Shigella flexneri* 2a Invaplex 50 vaccine. *Vaccine*. **29:** 7009-7019.

Roberts, F., Jennison, V. A. and Verma, N. K. (2005). The *Shigella flexneri* serotype Y vaccine candidate SFL124 originated from a serotype 2a background. *FEMS: Immunology and Medical Microbiology*. **45**: 285-289.

Sakaguchi, T., Kohler, H., Gu, X., McCormick, B. A. and Reinecker, H. C. (2002). *Shigella flexneri* regulates tight junction-associated proteins in human intestinal epithelial cells. Cell Microbiology **4**: 367-381

Sakellaris, H., Hannink, N. K., Rajakumar, K., Bulach, D., Hunt, M., Sasakawa C. and Adler, B. (2000). Curli loci of *Shigella* spp. *Infection and Immunity*. **68**: 3780-3783

Salam, M. A. (1998). Antimicrobial therapy for shigellosis: issues on antimicrobial resistance. *Japanese Journal of Medical Science and Biology.* **51**: S43-S62.

Samandari, T., Kotloff, K. L., Losonsky, G. L., Picking, W. D., Sansonetti, P. J., Levine, M. M. and Sztein, M. B. (2000). Production of IFN-gamma and IL-10 to Shigella invasins by mononuclear cells from volunteers orally inoculated with a Shiga toxin-deleted Shigella dysenteriae type 1 strain. *Journal of Immunology*. **164**: 2221-2232.

Sambrook, J. and Russell, W. D. (2001). Molecular Cloning: A Laboratory Manual (Third ed.). Cold Spring Harbor Laboratory, New York: *Cold Spring Harbor Laboratory Press*.

Sansonetti, P.J. (2001). Rupture, invasion and inflammatory destruction of the intestinal barrier by *Shigella*, making sense of prokaryote-eukaryote cross-talks. *FEMS Microbiology Letters*, **25**: 3-14.

Sansonetti, P. J. (2001). Shigellosis: from symptoms to molecular pathogenesis. *American Journal of Physiology*. **280**: 319-323.

Sasaki, T., Uchida, T. and Kurahashi, K. (1974). Glucosylation of O-antigen in *Salmonella* carrying epsilon 15 and epsilon 34 in phages. Journal Biological Chemistry. **249**: 761-772.

Schmidt, H. and M. Hensel. (2004). Pathogenicity islands in bacterial pathogenesis. *Clinical Microbiology Reviews*. **17**:14–56

Schroeder, G. N. and Hilbi, H. (2008). Molecular pathogenesis of Shigella spp.: Controlling host cell signaling, invasion, and death by type III secretion. *Clinical Microbiology Reviews* **21**: 134-156.

Schuch, R. and Maurelli, A. T. (1997). Virulence plasmid instability in *Shigella flexneri* 2a is induced by virulence gene expression. *Infection and Immunity*. **65**: 3686-3692.

Seal, R.P., Leighton, B.H. and Amara, S.G. (2000). A model for the topology of excitatory amino acid transporters determined by the extracellular accessibility of substituted cysteines. *Neuron*. **25**: 695-706.

Sellge, G., Magalhaes, J. G., Kornradt, C., Fritz, J. H., Salgao-Pabon, W., Eberi, G., Banderia, A., Di Santo, P. J. and Phailipon, A. (2010). Th17 cells are the dominant T cell subtype primed by Shigella flexneri mediating protective immunity. *Journal of Immunology*. **184**:2076–2085.

Sereny, B. (1957). Experimental keratoconjunctivitis Shigellosa. *Acta Microbiol Hungarica*, **4**: 367.

Shears, P. (1996). Shigella infections. *Annals of Tropical Medicine and Parisitology*. **90**: 105-114.

Skippington, E. and Ragan, M. A. (2011). Within-species lateral genetic transfer and the evolution of transcriptional regulation in *Escherichia coli* and *Shigella*. *BMC Genomics*. **12**. 532.

Simmons, D. A. and Ramanowska, E. (1987). Structure and biology of *Shigella flexneri* O antigens. Journal of Medical Microbiolgy. **23**: 289-302.

Skoudy, A., Mounier, J., Aruffo, A., Ohayon, H., Gounon, P., Sansonetti, P. and Tran Van Nhieu, G., (2000). CD44 binds to the *Shigella* IpaB protein and participates in bacterial invasion of epithelial cells. *Cell Microbiology*. **2**: 19-33.

Small, P., Blankenhorn, D., Welty, D., Zinser, E. and Slonczewski, J. L. (1994). Acid and base resistance in *Escherichia coli* and *Shigella flexneri*: role of *rpoS* and growth pH. *Journal Bacteriology*. **176**: 1729-1737.

Snider, J., Kittanakom, S., Curak, J. and Stagljar, I. (2010). Split-Ubiquitin Based Membrane Yeast Two-Hybrid (MYTH) System: A Powerful Tool For Identifying Protein-Protein Interactions. *Journal of Visualized Experiments*. **36**: 1698.

Snider, J., Kittanakom, S., Damjanovic, D., Curak, J., Wong, V. and Stagljar, I. (2010). Detecting interactions with membrane proteins using a membrane two-hybrid assay in yeast. *Nature Protocols*. **5**: 1281-1293.

Sonders, M.S., Amara, S.G. (1996). Channels in transporters. *Current Opinions in Neurobiology*. **6**: 294-302.

Sonnhammer, E. L., von Heijne, G. and Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proclaimed International Conference on Intellegent Systems for Molecular Biology*. **6**:175-182.

Sousa, M. A. B., Mendes, E. N., Apolonio, A. C. M., Farias L. D. M. and Magalhaes, P. P. (2010). Bacteriocin production by Shigella sonnei isolated from faeces of children with acute diarrhoea. *APMIS* **118**: 125-135.

Stagg, R. M., Cam, P. D. and Verma, N. K. (2008). Erratum: Identification of newly recognized serotype 1c as the most prevalent Shigella flexneri serotype in northern rural Vietnam. *Epidemiology and Infection*. **136**: 1141.

Stagg, R. M., Tang, S. S., Carlin, N. I. A., Talukder, K. A., Cam, P. D. and Verma, N. K. (2009). A novel glucosyltransferase involved in O-antigen modification of Shigella flexneri serotype 1c. *Journal of Bacteriology* **191**: 6612-6617.

Svennerholm, S. and Steele., D. (2004). Microbial-gut interactions in health and disease. Progress in enteric vaccine development. *Best Practice and Research Clinical Gastroenterology*. **18**: 421-445.

Sugiyama, J. E., Mahmoodian, S. and Jacobson, G. R. (1991). Membrane topology analysis of *Escherichia coli* mannitol permease by using a nested-deletion method to create *mtlA-phoA* fusions. *Proceedings of the National Academy of* Sciences. **88:** 9603-9607.

Sun, Q., Lan, R., Wang, Y., Wang, J.,, Luo, X.,, Zhang, S., Li, P., Wang,Y., Ye, C., Jing, H. and Xu, J. (2011). Genesis of a novel Shigella flexneri serotype by sequential infection of serotype-converting bacteriophages SfX and SfI. *BMC Microbiology*. **11**: 269.

Suzuki, Y., Hatagaki, K. and Oda, H. (1991). A hyperthermostable pullulanase produced by an extreme thermophile, Bacillus flavocaldarius KP 1228, and evidence for the proline theory of increasing protein thermostability. *Applied Microbiology and Biotechnology*. **34**: 707-714.

Talukder, K. A., Dutta, D. K., Safa, A., Ansaruzzaman, M., Hassan, F., Alam, K., Islam, K. M., Carlin, N. I., Nair, G. B. and Sack, D. A. (2001). Altering trends in the dominance of *Shigella flexneri* serotypes and emergence of serologically atypical *S. flexneri* strains in Dhaka, Bangladesh. *Journal of Clinical Microbiology*. **39:** 3757-3759.

Talukder, K. A., Islam, M. A., Dutta, D. K., Hassan, F., Sada, A., Nair., G. B. and Sack, D. A., (2002). Phenotypic and genotypic characterization of serologically atypical strains of Shigella flexneri type 4 isolated in Dhaka, Bangladesh. *Journal of Clinical Microbiology*. **40**:2490–2497.

Talukder, K. A., Islam, Z., Islam, M. A., Dutta, D. K., Safa, A., Ansaruzzaman, M., Faruque, A. S, Shahed, S. N., Nair, G. B. and Sack, D. A. (2003). Phenotypic and genotypic characterization of provisional serotype *Shigella flexneri* 1c and clonal relationships with 1a and 1b strains isolated in Bangladesh. *Journal of Clinical Microbiology*. **41:**110-117

Taylor, D., Echeverria, P., Pal, T., Sethabutr, O., Saiborisuth, S., Sricharmorn, S., Rowe, B. and Cross, J. (1986). The role of Shigella spp., enteroinvasive Escherichia coli, and other enteropathogens as causes of childhood dysentery in Thailand. *The Journal of Infectious Diseases* **153**: 1132-1138.

Thaminy, S., Auerbach, D., Arnoldo, A. and Stagljar. I. (2003). Identification of novel ErbB3-interacting factors using the split-ubiquitin membrane yeast two-hybrid system. Genome Research. **13**: 1744-1753.

Thanweer, F., Tahiliani, V., Korres, H. and Verma, N. K. (2008). Topology and identification of critical residues of the O-acetyltransferase of serotype-converting bacteriophage, SF6, of Shigella flexneri. *Biochemical and Biophysical Research Communications*. **375**: 581-585.

Tominaga, A., Mahmoud, M. A., Mukaihara, T. and Enomoto, M. (1994). Molecular characterization of intact, but cryptic, flagellin gene in genus *Shigella. Molecular Microbiology*. **12**: 277-285

Tran, V. (2010). Structural; and Function Analysis of GtrIC: a glucosyltransferase involved in the O-antigen modification of *Shigella flexneri* (Honours Thesis).

Tribble, D., Kaminski, R., Cantrell, J., Nelson, M., Porter, C., Baqar, S., Williams, C., Arora, R., Saunders, J., Ananthakrishnan, M., Sanders, J., Zaucha, G., Turbyfill, R. and E. Oaks. (2010). Safety and immunogenicity of a *Shigella flexneri* 2a Invaplex 50 intranasal vaccine in adult volunteers. *Vaccine* **28**: 6076-6085.

Turner, S. A., Luck, S. N., Sakellaris, H., Rajakumar, K. and Adler, B. (2001). Nested deletions of the SRL pathogenicity island of *Shigella flexneri* 2a. *Journal of Bacteriology*. **183:** 5535–5543.

Turner, S. A., Luck, S. N., Sakellaris, H., Rajakumar, K. and Adler, B.(2003). Molecular epidemiology of the SRL pathogenicity island. Antimicrobial Agents and Chemotherapy. **47:** 727–734.

Tusnady ,G.E. and Simon, I. (1998). Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *Journal of Molecular Biology.* **283:**489-506.

Ulmschneider, M. and M. Sansom. (2001). Amino acid distributions in integral membrane protein structures. *Biochimica et Biophysica Acta (BBA) – Biomembranes*. **1512**: 1-14.

Unligil, U.M. and Rini, J. M. (2000). Glycosyltransferase structure and mechanism. *Current Opinions in Structural Biology.* 10: 510-517.

van De Verg, L.L., Mallet, C.P., Collins, H.H., Larsen, T., Hammack, C. and Hale, T.L. (1995) Antibody and cytokine responses in a mouse pulmonary model of *Shigella flexneri* serotype 2a infection. *Infection and Immunity*, **63**:1947-1954.

Veenendaal, A. K., Hodgkinson, J. L., Schwarzer, L., Stabat, D., Zenk, S. F. and Blocker, A. J. The type III secretion system needle tip complex mediates host cell sensing and translocon insertion. *Molecular Microbiology*. **63**: 1719-1730.

Verma, N.K., Brandt, J.M., Verma, D.J. and Lindberg, A.A.: (1991). Molecular characterisation of the O-acetyl transferase gene of converting bacteriophage SF6 that adds group antigen 6 to *Shigella flexneri*. *Molecular Microbiology*. **5**: 71-75.

Verma, N. K., Verma, D. J., Huan, P. T. and Lindberg, A. A. (1993). Cloning and sequencing of the glucosyl transferase-encoding gene from converting bacteriophage X (SFX) of *Shigella flexneri*. *Gene.* **129**: 99-101.

von Heijne, G. (1992). Membrane protein structure prediction. Hydrophobicity 1 analysis and the positive-inside rule. *Journal of Molecular* Biology. **22**: 487-494.

von Seidlein, L., Kim, D. R., Ali, M., Lee, H., Wang, X., Thiem, V. D., Canh, D. G., Chaicumpa, W., Agtini, M. D., Hossain, A., Bhutta, Z. A., Mason, C., Sethabutr, O., Talukder, K., Nair, G. B., Deen, J. L., Kotloff., and Celemens, J. (2006). A multicentre study of Shigella diarrhoea in six Asian countries: disease burden, clinical manifestations, and microbiology. *PLoS Medicine.* **3**:e353.

Vu Nguyen, T., Le Van, P., Le Huy, C., Nguyen Gia, K. and Weintraub, A. (2006). Etiology and epidemiology of diarrhea in children in Hanoi,Vietnam. *International Journal of Infectious Disease* **10**: 298-308.

Wagner, S., Bader, M. L., Drew, D. and de Gier, J.W. (2006) Rationalizing membrane protein overexpression. Trends in Biotechnology **24**: 364–371.

Wagner, S., Klepsch, M. M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., Högbom, M., Van Wijk, K. J., Slotboom, D. J., Persson, J. O. and De Gier,J. W. (2008). Tuning Escherichia coli for membrane protein overexpression. *Proceedings of the National Academy of Sciences of the United States of America*. **105**: 14371-14376.

Wang, A. (2011). Functional analysis of GtrIc: a novel glucosyltransferase in serotype conversion in *Shigella flexneri* (Honours Thesis).

Wassef, J. S., Keren, D. F. and Mailloux, J. L. (1989). Role of M cells in initial antigen uptake and in ulcer formation in the rabbit intestinal loop model of shigellosis. Infection and Immunity **57**: 858-863.

Watarai, M., Funato, S. and Sasakawa, C. (1996). Interaction of Ipa proteins of *Shigella flexneri* with alpha5beta1 integrin promotes entry of the bacteria into mammalian cells. *Journal of Experimental Medicine* **183**: 991-999.

Way, S.S., Borczuk, A.C., Dominitz, R. and Goldberg M,B. (1998). An essential role for gamma interferon in innate resistance to *Shigella flexneri* infection. *Infection and Immunity*, **66**:1342-1348.

Wei, J., Goldberg, M. B., Burland, V., Venkatesan, M. M., Deng, W., Fournier, G., Mayhew, G. F., Plunkett, G 3rd., Rose, D.J., Darling, A., Perna, N. T., Payne, S. M., Runyen-Janecky, L. J., Zhou, S., Schwartz, D. C. and Blattner, F. R., (2003). Complete genome sequence and compartive genomics of Shigella flexneri serotype 2a strain 2457T. *Infection and Immunology*. **71**: 2775-2786.

Weinrauch, Y., Drujan, D., Shapiro, S.D., Weiss, J., Zychlinsky, A. (2002): Neutrophil elastase targets virulence factors of enterobacteria. *Nature*.**417**:91-4.

West, N.P., Sansonetti, P., Mounier, J., Exley, R.M., Parsot, C., Guadagnini, S., Prevost, M.C., Prochnicka-Chalufour, A., Delepierre, M., Tanguy, M. (2005). Optimization of virulence functions through glucosylation of Shigella LPS. *Science*. **307**:1313-1317.

Wolin, C.D. and Kaback, H.R. (2000). Thiol cross-linking of transmembrane domains IV and V in the lactose permease of *Escherichia coli. Biochemistry*. **39**: 6130-6135.

Williamson, M. P. (1994). The structure and function of proline-rich regions in proteins. *Biochemical Journal*. **297**: 249-260.

Wright, A. (1971). Mechanism of conversion of the salmonella O antigen by bacteriophageepsilon 34. *In Book Mechanism of conversion of the salmonella O antigen by bacteriophageepsilon* **105**: 927-36.

Wong, M. R., Reddy, V., Hanson, H., Johnson, K. M., Tsoi, B., Cokes, C., Gallagher, L., Lee, L., Plentsova, A., Dang, T., Krueger, A., Joyce, K. and Balter,S. (2010). Antimicrobial resistance trends of Shigella serotypes in New York City, 2006-2009. *Microbial Drug Resistance* **16**: 155-161.

Yaffe, M. B. (2005). X-ray crystallography and structural biology. *Critical Care Medicine*. **33**: S435-440.

Yang, F., Yang. J., Zhang, X., Chen, L., Jiang, Y., Yan, Y., Tang, X., Wang, J., Xiong, Z., Dong, J., Xue, Y., Zhu,Y., Xu, X., Sun, L., Chen, S., Nie,H., Peng, J., Xu, J., Wang, Y., Yuan, Z., Wen, Y., Yao, Z., Shen, Y., Qiang, B., Hou, Y., Yu, J. and Jin, Q., (2005) Genome dynamics and diversity of Shigella species, the etiologic agents of bacillary dysentery. *Nucleic Acids Research* **33**: 6445-6458.

Yang, J., Nie, H., Chen, L., Zhang, X., Yang, F., Xu, X., Zhu, Y., Yu J. and Jin Q. *(*2007). Revisiting the molecular evolutionary history of Shigella spp. *Journal of Molecular Evolution.* **64**: 71-79.

Ye C., Lan, R., Xia, S., Zhang, J., Sun, Q., Zhang, S., Jing, H., Wang, L., Li, Z., Zhou, Z., Zhao, A., Cui, Z., Cao, J., Jin, D., Huang, L., Wang, Y., Luo, X., Bai, X., Wang, Y., Wang, P., Xu, Q. and Xu,J . (2010). Emergence of a new multidrug-resistant serotype X variant in an epidemic clone of Shigella flexneri. *Journal of Clinical Microbiology.* **48**: 419-426.

Zhang Z, Jin, L., Champion, G., Seydel, K.B. and Stanley, S.L. Jr. (2001) Shigella infection in a SCID mouse-human intestinal xenograft model: role for neutrophils in containing bacterial dissemination in human intestine. *Infection and Immunity*. **69**:3240-3247.

Zhu, L., Lin, X., Zheng, X., Bu, X., Zhao, G., Xie, C., Zhang, J., Li, N., Feng, E, Wang, J., Jiang, Y., Huang, P and Wang, H. (2009). Global analysis of a plasmid-cured Shigella flexneri strain: New insights into the interaction between the chromosome and a virulence plasmid. *Journal of Proteome Research* **9**: 843-854.

Appendices

Appendix A

Buffers, Solutions and Media

All solutions were made up to volume with sterile MilliQ water and stored at room temperature unless otherwise stated.

Bacterial Growth Media

LB Broth

4g Tryptone

2g Yeast Extract

2g NaCl

Make up to 400 mL in MilliQ water and sterilize by autoclaving

LB Agar

As for LB broth, adding 15 g/L agar before autoclaving

IPTG / X-Gal Plates

Add 20µL each of X-Gal and IPTG solutions (20 mg/mL) and spread on plate

Plasmid DNA Isolation by Alkaline Lysis

Solution I

50 mM Glucose

25 mM Tris.HCl pH 8.0

10 mM EDTA pH 8.0

Sterilised by autoclaving. Stored at 4°C

Solution II

0.2M NaOH

1% SDS

Freshly prepared from sterile stock solutions (10% SDS, 3 M NaOH) before use

Solution III

60 mL 5 M CH_3COOK

11.5 mL Glacial acetic acid

28.5 mL Water

Prepared from sterile stock solutions and stored at 4°C

Agarose Gel Electrophoresis

5X TBE Buffer

54 g Tris Base

27.5 g Boric acid

20 mL 0.5 M EDTA pH 8.0

Make up to 1 L with water

Blue Loading Buffer

1 mg/mL Bromophenol blue

20% Glycerol

Dual Indicator Plates

Supplement LB agar with:

80 µg/ml X-phos

100 µg/mL Red Gal

1 mM IPTG

80 mM K_2HPO_4

(antibiotics for selection added to above mixture just before pouring plates)

β-Galactosidase Assay

Z buffer

0.06 M $Na_2HPO_4.7H_2O$

0.04 M $NaH_2PO_4.H_2O$

0.01 M KCl

0.001 M $MgSO_4$

0.27% (v/v) 2-mercaptoethanol

pH 7.0

(store at 4°C, make fresh)

Phosphate Buffer

0.06 M $Na_2HPO_4.7H_2O$

0.04 M $NaH_2PO_4.H_2O$

pH 7.0

SDS-PAGE

2x Sample Loading Buffer

10% (w/v) SDS

20% (v/v) glycerol

0.1% (w/v) bromophenol blue

0.5 M Tris-HCl (pH 6.8)

5% (v/v) 2-mercaptoethanol (add fresh before boiling)

10x SDS-PAGE Running Buffer

0.029 g/mL Tris Base

0.144 g/mL glycine

0.01% (w/v) SDS

12% Resolving gel solution

3.2 mL MilliQ Water

2.1 mL Bis Acrylamide

1.9 mL Tris HCl (1.5M pH8.8)

75 µL 10% SDS (w/v)

75 µL Ammonium Persulfate

3 µL Tetramethylethylenediamine

4% Stacking Gel

1.6 mL MilliQ Water

250 µL Bis Acrylamide

625 µL Tris HCl (0.5M pH6.8)

25 µL 10% SDS (w/v)

12.5 µL Ammonium Persulfate

2.5 µL Tetramethylethylenediamine

Silver Staining

Fixing Solution (100ml)

7.5 mL Glacial Acetic Acid

25 mL Propan-2-ol

Made up to 100 mL with MilliQ Water

Oxidizing Solution (100mL)

0.7 g Periodic Acid

7.5 mL Glacial Acetic Acid

Made up to 100 mL with MilliQ Water

Silver Staining Solution (100mL)

187 µL Sodium Hydroxide (10M)

1.3 mL Ammonium Hydroxide

0.67 g Silver Nitrate

Made up to 100 mL with MilliQ Water

Developer Solution (100mL)

22.2 µL Formaldehyde

5 mg Citric Acid

Made up to 100 mL with MilliQ Water

Coomassie Staining

Coomassie Staining Solution (1L)

0.25 g Comassie Brilliant Blue

75 mL Acetic Acid

500 mL Methanol

Made up to 1 L with MilliQ Water

Destaining Solution (1L)

400 mL Methanol

100 mL Acetic Acid

Made up to 1 L with MilliQ Water

Western Immunoblotting

1 x PBS (500 mL)

4 g NaCl

0.1 g KCl

0.72 g Na_2HPO_4

0.12 g KH_2PO_4

Made up to 500 mL with MilliQ Water

1 x PBS + 0.05% Tween20 (1 L)

8 g NaCl

0.2 g KCl

1.44 g Na_2HPO_4

0.24 g KH_2PO_4

0.5 mL Tween 20

Made up to 1 L with MilliQ water

Blocking Solution (50 mL)

2.5 g Skim Milk

50 mL 1 x PBS

Transfer Buffer (1 L)

6.05 g Tris base

28.82 g Glycine

1 g SDS

200 mL methanol

Made up to 1 L with MilliQ water

Southern Hybridization

20 x SSC(1L)

175.31 g NaCl (→3M)

88.23 g tri-Na citrate (→0.3M)

pH to 7

Make up to 1 L with MilliQ water and sterilize by autoclaving

Denaturing Solution (500 mL)

1.5 M NaCl

0.5 M NaOH

Make up to 500 mL with MilliQ water and sterilize by autoclaving

Neutralising solution (500 mL)

1.5 M NaCl (43.83g)

0.5 M Tris-HCl (pH = 7.2) (39.40g)

1 mM EDTA (0.146g)

Make up to 500 mL with MilliQ water and sterilize by autoclaving

0.25M HCl (400 mL)

10 mL of 10 M HCl stock

390 mL MilliQ water

10x Maleic acid buffer (500 m L)

0.1 M Maleic Acid (116.08 g / L)

0.15 M NaCl (87.66 g / L)

Adjust to pH 7,5 with 5 M NaOH

Make up to 500 mL with MilliQ water and sterilize by autoclaving

DIG detection buffer

0.1M NaCl

0.1 M Tris-HCl (pH 9.5)

DIG-Washing buffer

0.3 % (v/v) Tween-20 in DIG Maleic acid buffer

Appendix B

NCTC GtrIV alignments of wildtype serotype 4 strains against NCTC 8296

```
SFL1758          MAERIRYVFLAIFCALLCKESIDIWMMNNGDFDRAITPFMNGIRQFSHDGTLLYTLKENF 60
SFL1768          MAERIRYVFLAIFCALLCKESIDIWMMNNGDFDRAITPFMNGIRQFSHDGTLLYTLKENF 60
SFL1714          MAERIRYVFLAIFCALLCKESIDIWMMNNGDFDRAITPFMNGIRQFSHDGTLLYTLKENF 60
SFL1305          MAERIRYVFLAIFCALLCKESIDIWMMNNGDFDRAITPFMNGIRQFSHDGTLLYTLKENF 60
SFL1526          MAERIRYVFLAIFCALLCKESIDIWMMNNGDFDRAITPFMNGIRQFSHDGTLLYTLKENF 60
SFL2178          MAERIRYVFLAIFCALLCKESIDIWMMNNGDFDRAITPFMNGIRQFSHDGTLLYTLKENF 60
SFL1314          MAERIRYVFLAIFCALLCKESIDIWMMNNGDFDRAITPFMNGIRQFSHDGTLLYTLKENF 60
SFL2241          MAERIRYVFLAIFCALLCKESIDIWMMNNGDFDRAITPFMNGIRQFSHDGTLLYTLKENF 60
SFL1522          MAERIRYVFLAIFCALLCKESIDIWMMNNGDFDRAITPFMNGIRQFSHDGTLLYTLKENF 60
SFL1769          MAERIRYVFLAIFCALLCKESIDIWMMNNGDFDRAITPFMNGIRQFSHDGTLLYTLKENF 60
SFL1524          MAERIRYVFLAIFCALLCKESIDIWMMNNGDFDRAITPFMNGIRQFSHDGTLLYTLKENF 60
NCTC GtrIV       MAERIRYVFLAIFCALLCKESIDIWMMNNGDFDRAITPFMNGIRQFSHDGTLLYTLKENF 60
SFL1523          MAERIRYVFLAIFCALLCKESIDIWMMNNGDFDRAITPFMNGIRQFSHDGTLLYTLKENF 60
SFL2187          MAERIRYVFLAIFCALLCKESIDIWMMNNGDFDRAITPFMNGIRQFSHDGTLLYTLKENF 60
                 ************************************************************

SFL1758          SSIVNYEYKSSFSYILYLYAYVISIFTHTFDLRLWSSLSKIVYICALYKLFVEINNKKYS 120
SFL1768          SSIVNYEYKSSFSYILYLYAYVISIFTHTFDLRLWSSLSKIVYICALYKLFVEINNKKYS 120
SFL1714          SSIVNYEYKSSFSYILYLYAYVISIFTHTFDLRLWSSLSKIVYICALYKLFVEINNKKYS 120
SFL1305          SSIVNYEYKSSFSYILYLYAYVISIFTHTFDVRLWSSLSKIVYICALYKLFVEINNKKYS 120
SFL1526          SSIVNYEYKSSFSYILYLYAYVISIFTHTFDVRLWSSLSKIVYICALYKLFVEINNKKYS 120
SFL2178          SSIVNYEYKSSFSYILYLYAYVISIFTHTFDLRLWSSLSKIVYICALYKLFVEINNKKYS 120
SFL1314          SSIVNYEYKSSFSYILYLYAYVISIFTHTFDLRLWSSLSKIVYICALYKLFVEINNKKYS 120
SFL2241          SSIVNYEYKSSFSYILYLYAYVISIFTHTFDLRLWSSLSKIVYICALYKLFVEINNKKYS 120
SFL1522          SSIVNYEYKSSFSYILYLYAYVISIFTHTFDLRLWSSLSKIVYICALYKLFVEINNKKYS 120
SFL1769          SSIVNYEYKSSFSYILYLYAYVISIFTHTFDLRLWSSLSKIVYICALYKLFVEINNKKYS 120
SFL1524          SSIVNYEYKSSFSYILYLYAYVISIFTHTFDLRLWSSLSKIVYICALYKLFVEINNKKYS 120
NCTC GtrIV       SSIVNYEYKSSFSYILYLYAYVISIFTHTFDLRLWSSLSKIVYICALYKLFVEINNKKYS 120
SFL1523          SSIVNYEYKSSFSYILYLYAYVISIFTHTFDLRLWSSLSKIVYICALYKLFVEINNKKYS 120
SFL2187          SSIVNYEYKSSFSYILYLYAYVISIFTHTFDLRLWSSLSKIVYICALYKLFVEINNKKYS 120
                 ********************************:***************************

SFL1758          YLLFPVSALLLFTPSILSQFTAFYQEQIVITLLPVITYIILKKEISRFDKVLLVACAALI 180
SFL1768          YLLFPVSALLLFTPSILSQFTAFYQEQIVITLLPVITYIILKKEISRFDKVLLVACAALI 180
SFL1714          YLLFPVSALLLFTPSILSQFTAFYQEQIVITLLPVITYIILKKEISRFDKVLLVACAALI 180
SFL1305          YLLFPVSALLLFTPSILSQFTAFYQEQIVITLLPVITYIILKKEISRFDKVLLVACAALI 180
SFL1526          YLLFPVSALLLFTPSILSQFTAFYQEQIVITLLPVITYIILKKEISRFDKVLLVACAALI 180
SFL2178          YLLFPVSALLLFTPSILSQFTAFYQEQIVITLLPVITYIILKKEISRFDKVLLVACAALI 180
SFL1314          YLLFPVSALLLFTPSILSQFTAFYQEQIVITLLPVITYIILKKEISRFDKVLLVACAALI 180
SFL2241          YLLFPVSALLLFTPSILSQFTAFYQEQIVITLLPVITYIILKKEISRFDKVLLVACAALI 180
SFL1522          YLLFPVSALLLFTPSILSQFTAFYQDQIVITLLPVITYIILKKEISRFDKVLLVACAALI 180
SFL1769          YLLFPVSALLLFTPSILSQFTAFYQDQIVITLLPVITYIILKKEISRFDKVLLVACAALI 180
SFL1524          YLLFPVSALLLFTPSILSQFTAFYQDQIVITLLPVITYIILKKEISRFDKVLLVACAALI 180
NCTC GtrIV       YLLFPVSALLLFTPSILSQFTAFYQDQIVITLLPVITYIILKKEISRFDKVLLVACAALI 180
SFL1523          YLLFPVSALLLFTPSILSQFTAFYQDQIVITLLPVITYIILKKEISRFDKVLLVACAALI 180
SFL2187          YLLFPVSALLLFTPSILSQFTAFYQDQIVITLLPGITYVILKKEISRFDRVLLVACAALI 180
                 ***************************:******** ***:**********:**********
```

```
SFL1758          ATSKSQFFYIPILISLISIIFIDIRDKKFHLAMVIATIIALSFSFFSPSATKNNSYHSLY 240
SFL1768          ATSKSQFFYIPILISLISIIFIDIRDKKFHLAMVIATIIALSFSFFSPSATKNNSYHSLY 240
SFL1714          ATSKSQFFYIPILISLISIIFIDIRDKKFHLAMVIATIIALSFSFFSPSATKNNSYHSLY 240
SFL1305          ATSKSQFFYIPILISLISIIFIDIRDKKFHLAMVIATIIALSFSFFSPSATKYNSYHSLY 240
SFL1526          ATSKSQFFYIPILISLISIIFIDIRDKKFHLAMVIATIIALSFSFFSPSATKYNSYHSLY 240
SFL2178          ATSKSQFFYIPILISLISIIFIDIRDKKFHLAMVIATIIALSFSFFSPSATKYNSYHSLY 240
SFL1314          ATSKSQFFYIPILISLISIIFIDIRDKKFHLAMVIATIIALSFSFFSPSATKYNSYHSLY 240
SFL2241          ATSKSQFFYIPILISLISIIFIDIRDKKFHLAMVIATIIALSFSFFSPSATKNNSYHSLY 240
SFL1522          ATSKSQFFYIPILISLISIIFIDIRDKKFHLAMVIATIIALSFSFFSPSATKYNSYHSLY 240
SFL1769          ATSKSQFFYIPILISLISIIFIDIRDKKFHLAMVIATIIALSFSFFSPSATKYNSYHSLY 240
SFL1524          ATSKSQFFYIPILISLISIIFIDIRDKKFHLAMVIATIIALSFSFFSPSATKYNSYHSLY 240
NCTC GtrIV       ATSKSQFFYIPILISLISIIFIDIRDKKFHLAMVIATIIALSFSFFSPSATKNNSYHSLY 240
SFL1523          ATSKSQFFYIPILISLISIIFIDIRDKKFHLAMVIATIIALSFSFFSPSATKNNSYHSLY 240
SFL2187          ATSKSQFFYIPILISLISIIFIDIRDKKFHLAMVIATIIALSFSFFSPSATKNNSYHSLY 240
                 ************************************************** *******

SFL1758          FGTLLYNKSNGKENPSWAVDECIGVDAWGNKFDLDKGAVTTEDAGACFAKNKERGLKDSL 300
SFL1768          FGTLLYNKSNGKENPSWAVDECIGVDAWGNKFDLDKGAVTTEDAGACFAKNKERGLKDSL 300
SFL1714          FGTLLYNKSNGKENPSWAVDECIGVDAWGNKFDLDKGAVTTEDAGACFAKNKERGLKDSL 300
SFL1305          FGTLLYNKSNGKENPSWAVDECIGVDAWGNKFDLDKGAVTTEDAGACFAKNKERGLKDSL 300
SFL1526          FGTLLYNKSNGKENPSWAVDECIGVDAWGNKFDLDKGAVTTEDAGACFAKNKERGLKDSL 300
SFL2178          FGTLLYNKSNGKENPSWAVDECIGVDAWGNKFDLDKGAVTTEDAGACFAKNKERGLKDSL 300
SFL1314          FGTLLYNKSNGKKNPSWAVDECIGVDAWGNKFDLDKGAVTTEDAGACFAKNKERGLKDSL 300
SFL2241          FGTLLYNKSNGKKNPSWAVDECIGVDAWGNKFDLDKGAVTTEDAGACFAKNKERGLKDSL 300
SFL1522          FGTLLYNKSNGKENPSWAVDECIGVDAWGNKFDLDKGAVTTKDAGACFAKNKERGLKDSL 300
SFL1769          FGTLLYNKSNGKENPSWAVDECIGVDAWGNKFDLDKGAVTTKDAGACFAKNKERGLKDSL 300
SFL1524          FGTLLYNKSNGKENPSWAVDECIGVDAWGNKFDLDKGAVTTEDAGACFAKNKERGLKDSL 300
NCTC GtrIV       FGTLLYNKSNGKENPSWAVDECIGVDAWGNKFDLDKGAVTTEDAGACFAKNKERGLKDSL 300
SFL1523          FGTLLYNKSNGKENPSWAVDECIGVDAWGNKFDLDKGAVTTEDAGACFAKNKERGLKDSL 300
SFL2187          FGTLLYNKSNGKENPSWAVDECIGVDAWGNKFDLDKGAVTTEDAGACFAKNKERGLKDSL 300
                 ************:*****************************:******************

SFL1758          LLMLKQPSMFLLLPFDSGVRTQLTEDYFHVFKENKLIISKSVFLNEVQKIKDSALSTVRI 360
SFL1768          LLMLKQPSMFLLLPFDSGVRTQLTEDYFHVFKENKLIISKSVFLNEVQKIKDSALSTVRI 360
SFL1714          LLMLKQPSMFLLLPFDSGVRTQLTEDYFHVFKENKLIISKSVFLNEVQKIKDSALSTVRI 360
SFL1305          LLMLKQPSMFLLLPFDSGVRTQLTEDYFHVFKENKLIISKSVFLNEVQKIKDSALSTVRI 360
SFL1526          LLMLKQPSMFLLLPFDSGVRTQLTEDYFHVFKENKLIISKSVFLNEVQKIKDSALSTVRI 360
SFL2178          LLMLKQPSMFLLLPFDSGVRTQLTEDYFHVFKENKLIISKSVFLNEVQKIKDSALSTVRI 360
SFL1314          LLMLKQPSMFLLLPFDSGVRTQLTEDYFHVFKENKLIISKSVFLNEVQKIKDSALSTVRI 360
SFL2241          LLMLKQPSMFLLLPFDSGVRTQLTEDYFHVFKENKLIISKSVFLNEVQKIKDSALSTVRI 360
SFL1522          LLMLKQPSMFLLLPFDSGVRTQLTEDYFHVFKENKLIISKSVFLNEVQKIKDSALSTVRI 360
SFL1769          LLMLKQPSMFLLLPFDSGVRTQLTEDYFHVFKENKLIISKSVFLNEVQKIKDSALSTVRI 360
SFL1524          LLMLKQPSMFLLLPFDSGVRTQLTEDYFHVFKENKLIISKSVFLNEVQKIKDSALSTVRI 360
NCTC GtrIV       LLMLKQPSMFLLLPFDSGVRTQLTEDYFHVFKENKLIISKSVFLNEVQKIKDSALSTVRI 360
SFL1523          LLMLKQPSMFLLLPFDSGVRTQLTEDYFHVFKENKLIISKSVFLNEVQKIKDSALSTVRI 360
SFL2187          LLMLKQPSMFLLLPFDSGVRTQLTEDYFHVFKENKLIISKSVFLNEVQKIKDSALSTVRI 360
                 ************************************************************
```

```
SFL1758          PVAFIAMIASLIFIRKKYSVAIFTLSAFSISQFYISFIGEGYRDLDKHLFAMNFSFDLMI 420
SFL1768          PVAFIAMIASLIFIRKKYSVAIFTLSAFSISQFYISFIGEGYRDLDKHLFAMNFSFDLMI 420
SFL1714          PVAFIAMIASLIFIRKKYSVAIFTLSAFSISQFYISFIGEGYRDLDKHLFAMNFSFDLMI 420
SFL1305          PVAFIAMIASLIFIRKKYSVAIFTLSAFSISQFYISFIGEGYRDLDKHLFAMNFSFDLMI 420
SFL1526          PVAFIAMIASLIFIRKKYSVAIFTLSAFSISQFYISFIGEGYRDLDKHLFAMNFSFDLMI 420
SFL2178          PVAFIAMIASLIFIRKKYSVAIFTLSAFSISQFYISFIGEGYRDLDKHLFAMNFSFDLMI 420
SFL1314          PVAFIAMIASLIFIRKKYSVAIFTLSAFSISQFYISFIGEGYRDLDKHLFAMNFSFDLMI 420
SFL2241          PVAFIAMIASLIFIRKKYSVAIFTLSAFSISQFYISFIGEGYRDLDKHLFAMNFSFDLMI 420
SFL1522          PVAFIAMIASLIFIRKKYSVAIFTLSAFSISQFYISFIGEGYRDLDKHLFAMNFSFDLMI 420
SFL1769          PVAFIAMIASLIFIRKKYSVAIFTLSAFSISQFYISFIGEGYRDLDKHLFAMNFSFDLMI 420
SFL1524          PVAFIAMIASLIFIRKKYSVAIFTLSAFSISQFYISFIGEGYRDLDKHLFAMNFSFDLMI 420
NCTC GtrIV       PVAFIAMIASLIFIRKKYSVAIFTLSAFSISQFYISFIGEGYRDLDKHLFAMNFSFDLMI 420
SFL1523          PVAFIAMIASLIFIRKKYSVAIFTLSAFSISQFYISFIGEGYRDLDKHLFAMNFSFDLMI 420
SFL2187          PVAFIAMIASLIFIRKKYSVAIFTLSAFSISQFYISFIGEGYRDLDKHLFAMNFSFDLMI 420
                 ************************************************************

SFL1758          FIVLSIILDKVASGIYK 437
SFL1768          FIVLSIILDKVASGIYK 437
SFL1714          FIVLSIILDKVASGIYK 437
SFL1305          FIVLSIILDKVASGIYK 437
SFL1526          FIVLSIILDKVASGIYK 437
SFL2178          FIVLSIILDKVASGIYK 437
SFL1314          FIVLSIILDKVASGIYK 437
SFL2241          FIVLSIILDKVASGIYK 437
SFL1522          FIVLSIILDKVASGIYK 437
SFL1769          FIVLSIILDKVASGIYK 437
SFL1524          FIVLSIILDKVASGIYK 437
NCTC GtrIV       FIVLSIILDKVASGIYK 437
SFL1523          FIVLSIILDKVASRIYK 437
SFL2187          FIVLSIILDKVASGIYK 437
                 ************* ***
```

The NCTC GtrIV sequence is highlighted in grey. All mutations are highlighted in yellow.